

Smith, Mona

79815

From: ~~Schreiber, David~~
Sent: Monday, November 04, 2002 4:41 PM
To: Smith, Mona
Cc: Nashed, Nashaat
Subject: 09/724,8746

AV 1652

10001
mailbox
please

Mona,

Please run the following fragments in the rush search for this case:

Seq ID no. 2, residues 16975-17065
residues 18300-18400
residues 19950-20050
residues 20490-20590
residues 21350-21450
residues 22750-22850
residues 24300-24400
residues 24950-25050
residues 25850-25950
residues 27200-27300
residues 28200-28300
residues 29800-29900
residues 30550-30650
residues 31170-31270
residues 32070-32170
residues 33500-33600
residues 34350-34450
residues 36150-36250
residues 36940-37040
residues 37570-37670

Point of Contact:
Mona Smith
Technical Information Specialist
CM1 6A01
Tel: 308 3278

David Schreiber, Ph.D.
Scientific and Technical Information Center
Biotech/Chem Library
CM1-6A03
703-308-4292

Searcher: M. Smith
Phone:
Location:
Date Picked Up: 11/5/02
Date Completed: 11/8/02
Searcher Prep/Review: 15
Clerical:
Online Time 25

NA sequences: 20
AA sequences:
Structures:
Bibliographic
Litigation:
Full text:
Patent Family
Other:

Vendor Cost:
STN:
Dialog:
Orbit/Questel:
Lexis/Nexis:
Sequence Sys:
WWW/internet:
Other:

Point of Contact
Mona Smith
Technical Information Specialist
CMJ 8A01
Tel 308 3038

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 1197.62 Seconds
(without alignments)
1365.823 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101
Sequence: 1 aggtctcgcaggtctgcgc.....tgggcctagagctcgcgaac 101

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthm: *
3: em_estln: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_estl: *
10: gb_est2: *
11: gb_hlc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rnd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	36.2	35.8	620	17	BH385803 AG-ND-138
2	36.2	35.8	629	17	BH372094 AG-ND-138
3	35.2	34.9	419	10	AW922099 LG1_236_G
4	35.2	34.9	568	10	BE599451 P11_87_F0
5	34.2	33.9	465	10	BE595103 P11_45_F0
6	33.6	33.3	219	10	AW679204 WSI_23_G0

7	33.6	33.3	384	10	BE363648	BE363648 WSI_64_G1
8	33.6	33.3	384	14	BO283127	BO283127 WHE3086_D
9	33.6	33.3	401	10	AW287455	AW287455 LG1_228_B
10	33.6	33.3	402	10	BE363414	BE363414 WSI_62_F0
11	33.6	33.3	404	10	AW564133	AW564133 LG1_282_D
12	33.6	33.3	434	10	BE363488	BE363488 WSI_63_G1
13	33.6	33.3	440	10	AW746099	AW746099 WSI_39_H0
14	33.6	33.3	453	10	AW745893	AW745893 WSI_38_C0
15	33.6	33.3	453	10	BE366824	BE366824 P11_41_A0
16	33.6	33.3	456	13	BC933006	BC933006 WSI_3_A04
17	33.6	33.3	456	13	BC933007	BC933007 WSI_3_A06
18	33.6	33.3	460	10	BE593071	BE593071 WSI_97_B0
19	33.6	33.3	464	10	BE594219	BE594219 WSI_103_G
20	33.6	33.3	467	10	AW747205	AW747205 WSI_66_C0
21	33.6	33.3	468	10	AW678238	AW678238 WSI_14_H0
22	33.6	33.3	468	10	BE597494	BE597494 P11_70_C0
23	33.6	33.3	468	10	BE599472	BE599472 P11_88_A0
24	33.6	33.3	473	10	AW746220	AW746220 WSI_40_D0
25	33.6	33.3	474	10	BE366769	BE366769 P11_41_A0
26	33.6	33.3	475	10	AW677810	AW677810 WSI_11_B0
27	33.6	33.3	475	10	BE364060	BE364060 P11_11_C1
28	33.6	33.3	476	10	AW746140	AW746140 WSI_39_H1
29	33.6	33.3	476	13	BM318387	BM318387 P11_13_A0
30	33.6	33.3	480	13	BM323992	BM323992 P1C1_30_A
31	33.6	33.3	490	13	BM328506	BM328506 P1C1_30_A
32	33.6	33.3	492	10	BE592512	BE592512 WSI_94_A0
33	33.6	33.3	503	10	BE594053	BE594053 WSI_101_F
34	33.6	33.3	504	10	BE593619	BE593619 WSI_98_H0
35	33.6	33.3	511	10	AW745041	AW745041 LG1_386_A
36	33.6	33.3	511	10	AW746104	AW746104 WSI_39_H1
37	33.6	33.3	524	10	AW745368	AW745368 WSI_34_A0
38	33.6	33.3	525	10	AW284481	AW284481 LG1_284_A
39	33.6	33.3	525	10	AW747678	AW747678 WSI_75_B0
40	33.6	33.3	527	10	BE599116	BE599116 P11_85_E0
41	33.6	33.3	527	13	BM326393	BM326393 P1C1_56_E
42	33.6	33.3	528	10	AW745988	AW745988 WSI_38_C0
43	33.6	33.3	532	10	AW746329	AW746329 WSI_40_D0
44	33.6	33.3	533	10	AW680220	AW680220 WSI_50_H0
45	33.6	33.3	538	10	AW746131	AW746131 WSI_39_H0

ALIGNMENTS

RESULT 1
BH385803
LOCUS
DEFINITION AG-ND-13813.TF.1 ND-TAM Anopheles gambiae genomic clone AG-ND-13813
, DNA sequence.
ACCESSION BH385803
VERSION BH385803.1 GI:17331945
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae
Anopheles malaria mosquito.
African malaria mosquito.
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Amophel.
REFERENCE
1 (bases 1 to 620)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
AUTHORS
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNLS
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..620

/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_1lb="AG-ND-13813"
/clone_1lb="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 112 a 211 c 200 g 97 t
ORIGIN

Query Match 35.8%; Score 36.2; DB 17; Length 620;

Best Local Similarity 65.4%; Pred. No. 6.1;

Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 8 GCAGGTCGCGCCGCTCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCTGGG 67

Db 276 GAAGCTGCTGCGCCGCTACGACGCGAAGCGGCTGTGCCCCCATGGGCTGGC 335

QY 68 AATGACTCGCTGATGGGCT 88

Db 336 CATGACGCTCTGCACAGCT 356

RESULT 2
BH372094 629 bp DNA linear GSS 10-DEC-2001

LOCUS AG-ND-13813.TF ND-TAM Anopheles gambiae genomic clone AG-ND-13813,
DEFINITION DNA sequence.

ACCESSION BH372094 GI:17318219

VERSION BH372094.1 GI:17318219

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

Anopheles.

1 (bases 1 to 629)

REFERENCE
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)

COMMENT
Other_GSSs: AG-ND-13813.TR
Contact: Brendan J Loftus

Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seg primer: M13 For

Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..629

/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_1lb="AG-ND-13813"
/clone_1lb="ND-TAM"

BASE COUNT 110 a /note="Vector: pECBAC1; Site_1: HindIII"
209 c 206 g 104 t
ORIGIN

Query Match 35.8%; Score 36.2; DB 17; Length 629;

Best Local Similarity 65.4%; Pred. No. 6.1;

Matches 53; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 8 GCAGGTCGCGCCGCTCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCTGGG 67

Db 276 GAAGCTGCTGCGCCGCTACGACGCGAAGCGGCTGTGCCCCCATGGGCTGGC 335

QY 68 AATGACTCGCTGATGGGCT 88

Db 336 CATGACGCTCTGCACAGCT 356

RESULT 3
AW922099 419 bp mRNA linear EST 19-JUL-2000

LOCUS LG1.236.G05.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.

ACCESSION AW922099 GI:8087924

VERSION AW922099.1 GI:8087924

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogonae; Sorghum.

1 (bases 1 to 419)

REFERENCE
AUTHORS Cordomier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)

COMMENT
Contact: Cordomier-Pratt MM

Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

Fax: 706 583 0210

Email: mmp@pratt.edu

Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV

High quality sequence stop: 208

POLYA-No.

FEATURES
source
Location/Qualifiers
1..419

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_1lb="Light Grown 1 (LG1)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda zap; Site_1: XhoI; Site_2: EcoRI

; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 76 a 130 c 149 g 64 t

ORIGIN

Query Match 34.9%; Score 35.2; DB 10; Length 419;

Best Local Similarity 65.0%; Pred. No. 10;

Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 22 TCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCTGGGAATGACTCGCTGA 81

Db 317 TCCCGAAGCAAGCTGAGTGCCTCCGCTCAGAGCCTGGGAATGACTCGCTGA 376

QY 82 TGGGCTAGAGCTGCGCAAC 101

Db 377 AGAGGCACTACTGGGCTAC 396

RESULT 4
BE599451
LOCUS
DEFINITION BE599451 568 bp mRNA linear EST 18-AUG-2000
P11_87_F08.b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION BE599451
VERSION BE599451.1 GI:9854524
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 568)
Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp Pratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 490
POLYA-No.

FEATURES
source
1..568
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
Vector: pBluescript II from lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
cultivar) were infected with pathogen (isolate FRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

BASE COUNT 103 a 175 c 194 g 96 t
ORIGIN

Query Match 34.9%; Score 35.2; DB 10; Length 568;
Best Local Similarity 65.0%; Pred. No. 11;
Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 22 TCCCGAAGGCAAGCTCGATGCGCGCTCAGAGAGCTGGAGATGACTGCTGA 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 285 TCCCGAGAGCCCAAGCTCCCGCGCGCGCGAGGCGAGCGATGAGGTGACGCGGTGG 344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 82 TGGGGCTAGAGCTGCGCAAC 101
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 AGAGGACTACTGCTCTAC 364
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
BE595103
LOCUS
DEFINITION BE595103 465 bp mRNA linear EST 18-AUG-2000
P11_45_F08.b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,

mRNA sequence.
ACCESSION BE595103
VERSION BE595103.1 GI:9850176
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 465)
Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp Pratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 342
POLYA-No.

FEATURES
source
1..465
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
Vector: pBluescript II from lambda Zap II; Site_1: XhoI;
Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
cultivar) were infected with pathogen (isolate FRM421 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants resume susceptibility
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

BASE COUNT 79 a 145 c 169 g 70 t 2 others
ORIGIN

Query Match 33.9%; Score 34.2; DB 10; Length 465;
Best Local Similarity 63.8%; Pred. No. 19;
Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 22 TCCCGAAGGCAAGCTCGATGCGCGCTCAGAGAGCTGGAGATGACTGCTGA 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 TCCCGAGAGCCCAAGCTCCCGCGCGCGCGAGGCGAGCGATGAGGTGACGCGGTGG 366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 82 TGGGGCTAGAGCTGCGCAAC 101
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 AGAGGACTACTGCTCTAC 386
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AM679204
LOCUS
DEFINITION AM679204 219 bp mRNA linear EST 19-JUL-2000
WS1_23_G01.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION AM679204
VERSION AM679204.1 GI:7552958
KEYWORDS EST.

SOURCE	sorghum.
ORGANISM	Sorghum bicolor
REFERENCE AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum. 1 (bases 1 to 219) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H. An EST database from Sorghum: water-stressed plants unpublished (2000)
TITLE JOURNAL COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20 Seq primer: JEN REV High quality sequence stop: 218 POLYA-No.
FEATURES source	Location/Qualifiers 1..219 <code>/organism="Sorghum bicolor"</code> <code>/db_xref="taxon:4558"</code> <code>/clone_id="Water-stressed 1 (WSI)"</code> <code>/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: lambda Zap I; Site:1: XhoI; Site:2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap IX. Clones to be sequenced were prepared by mass excision."</code> BASE COUNT 35 a 65 c 88 g 31 t ORIGIN
Query Match	33.3%; Score 33.6; DB 10; Length 219; Best local similarity 63.8% ; Pred. No. 23;
Matches	51: Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY	22 TCCCGAGGAACGTGCAGCTGGATGCCCGCCCTCACGACCCTGGAAATGCACTGCTGA 81 Db 93 TCCCAGAGCCACAAGTCGCGCGCGCGCGCGCGAGGCGAGAGAGGTTGAGCGCGGTG 152
OY	82 TGCGGCTAGACTGCCAAC 101 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
DB	153 AGAGGACTTACTTGCTTAC 172
RESULT 7 BE363648 LOCUS DEFINITION	BE363648 384 bp mRNA linear EST 20-JUN-2000 WSI_64.G12.g1.A002 Water-stressed 1 (WSI) Sorghum bicolor CDNA, mRNA sequence.
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	BE363648 BE363648.1 GI:305205 EST. Sorghum. Sorghum bicolour Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum. 1 (bases 1 to 384) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H. An EST database from Sorghum: water-stressed plants unpublished (2000) Contact: Cordonier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210

Email: mmprratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolYTmIX
High quality sequence start: 5
High quality sequence stop: 362
POLYA=yes

FEATURES
SOURCE Location/Qualifiers

 1..384
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Water-stressed l (WSl)"
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI;
 Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 74 a 114 c 129 g 67 t

ORIGIN

Oy Query Match 33.3%; Score 33.6; DB 10; Length 384;
Best Local Similarity 63.8%; Pred. No. 27;
Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 22 TCCCGAGGCACAGCTCGACGTGGGCCGCTCACAGCCTGGGAATGCACTTCGCGTA 81
 ||||| + ||||| | ||||| ||||| | ||||| ||||| ||||| ||||| |||||
Db 95 TCCCCGAGCCCACAAGCTTCCCGCGGCGCGCGGAGCGAGCAGAGAGGGTGAGCGCGGTGG 154
 ||||| + ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY 82 TGGGGCTTAGACTCGCGCAAC 101
 ||| | ||| | ||| | |||
Db 155 AGAGGGACTACTCTGCTCATC 174

RESULT 8
BOZ83127 384 bp mRNA linear EST I3-MAY-2002

LOCUS BOZ83127

DEFINITION WHR3086_D1L_H22S wheat cold-stressed seedling subtracted CDNA library Trilicium aestivum cdna clone WHR3086_D1L_H22, mRNA sequence.

ACCSSION BOZ83127

VERSION BOZ83127

KEYWORDS BQ283127.1 GI:20552527

SOURCE EST.

ORGANISM bread wheat.
Trilicium aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Trilicum.
1 (bases 1 to 384)
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Crossman,C., Fenton,R.D., Lazzo,G.R., Nguyen,H.T., Pham,J., Rausch,C.J., Wilson,C., Woo,J. and Zhang,D.
.The structure and function of the expressed portion of the wheat genomes - Cold-stressed seedling subtracted CDNA library unpublished (2002)

JOURNAL COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES
SOURCE Location/Qualifiers

 1..384
 /organism="Trilicum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone_lib="WHR3086 D1L H22"
 /clone_wbp="wheat cold-stressed seedling subtracted CDNA library"

[illegible]

QY 82 TGGGCTAGAGCTGCCAAC 101
 Db 152 AGAGGAGTACTGTCTCTAC 171

RESULT 11

LOCUS AM564133 404 bp mRNA EST 19-JUL-2000
 DEFINITION LG1_282.D02.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
 ACCESSION AM564133
 VERSION AM564133.1 GI:7218011
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 404)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
 An EST database from Sorghum: light-grown seedlings
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.

REFERENCE 1 (bases 1 to 404)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
 An EST database from Sorghum: light-grown seedlings
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.

Seq primer: JEN REV

High quality sequence stop: 403
 POLYA-No.

FEATURES

source

Location/Qualifiers

1..404

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Light Grown 1 (LG1)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse)
 seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
 ; The library was made from poly-A RNA in the cloning
 vector Lambda Zap II. Clones to be sequenced were
 prepared by mass excision."

BASE COUNT 71 a 127 c 144 g 62 t

ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 404;
 Best Local Similarity 63.8%; Pred. No. 27;
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 22 TCCCGGAGGCAAGCTGCAGTGGATGCCGCTCACAGACCTGGGAATGACTCGCTGA 81

Db 316 TCCCGGAGGCAAGCTGCAGTGGATGCCGCTCACAGACCTGGGAATGACTCGCTGA 375

QY 82 TGGGCTAGAGCTGCCAAC 101

Db 376 AGAGGAGTACTGTCTCTAC 395

RESULT 12

LOCUS BE363488 434 bp mRNA EST 20-JUL-2000
 DEFINITION WS1_63_G11.G1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BE363488
 VERSION BE363488.1 GI:9305045
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE 1 (bases 1 to 434)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: water-stressed plants
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.

Seq primer: PolyTMix
 High quality sequence start: 21
 High quality sequence stop: 434
 POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..434

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Water-stressed 1 (WS1)"

/note="Organ: Mix of 5-week old plants on days 7 & 8 after
 water was withheld; Vector: Lambda Zap; Site_1: XhoI;
 Site_2: EcoRI; The library was made from poly-A RNA in the
 cloning vector Lambda Zap II. Clones to be sequenced were
 prepared by mass excision."

BASE COUNT 73 a 141 c 157 g 63 t

ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 434;
 Best Local Similarity 63.8%; Pred. No. 28;
 Matches 51; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 22 TCCCGGAGGCAAGCTGCAGTGGATGCCGCTCACAGACCTGGGAATGACTCGCTGA 81

Db 198 TCCCGGAGGCAAGCTGCAGTGGATGCCGCTCACAGACCTGGGAATGACTCGCTGA 257

QY 82 TGGGCTAGAGCTGCCAAC 101

Db 258 AGAGGAGTACTGTCTCTAC 277

RESULT 13

LOCUS AW746099 440 bp mRNA EST 19-JUL-2000
 DEFINITION WS1_39_H04.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION AW746099
 VERSION AW746099.1 GI:7659837
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 440)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: water-stressed plants
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@prattuga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence

	Matches	51;	Conservative	0;	Mismatches	29;	Indels	0;	Gaps	0;
QY	22	TCCCGAAGGCAAGCTGAGGTGATGCCCGCTCACGAGCCTGGGAATGACTCGCTGA	81							
Db	15	TCCCGAAGGCAAGCTGAGGTGATGCCCGCTCACGAGCCTGGGAATGACTCGCTGA	74							
QY	82	TGGGCTAGAGCTGCGCAC	101							
Db	75	AGAGGACTACTGTCTTAC	94							

Search completed: November 6, 2002, 15:52:51
 Job time : 1200.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 : Search time 27.125 Seconds
(without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101
Sequence: 1 agctcgcaggtgctgcgc.....tggggctagagctgcgcac 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 16657418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications -NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.4	30.1	3209	10	US-09-925-301-474
2	29.8	29.5	11220	10	US-09-861-289-32
3	29.8	29.5	36778	10	US-09-861-289-5
4	29	28.7	1007	10	US-09-764-898-58
5	28.9	28.7	15872	10	US-09-861-289-1
6	28.8	28.5	290	10	US-09-294-0938-5935
7	28.8	28.5	2638	10	US-09-954-456-1955
8	28	27.7	2125	10	US-09-880-107-2412
9	28	27.7	4149	12	US-10-044-090-104
10	28	27.7	5086	10	US-09-880-107-3947
11	28	27.7	5145	10	US-09-925-299-206
12	28	27.7	5416	10	US-09-954-456-786
13	28	27.7	5432	12	US-09-880-107-2094
14	28	27.3	245	10	US-10-044-090-32
15	27.6	27.3	999	10	US-09-884-441-305
16	27.6	27.3	1482	10	US-09-815-242-7966
17	27.2	26.9	1395	10	US-09-815-242-7954
18	27.2	26.9	2172	10	US-09-815-242-8008
19	27.2	26.9	2172	10	US-09-815-242-4038

c	20	27	26.7	419	10	US-09-960-352-14518	Sequence 14518, A
c	21	27	26.7	456	10	US-09-864-761-41	Sequence 41, Appl
c	22	27	26.7	642	10	US-09-864-761-16883	Sequence 16883, A
c	23	27	26.7	2427	10	US-09-254-783A-2	Sequence 2, Appl
c	24	27	26.7	2427	12	US-10-152-058-2	Sequence 26, Appl
c	25	27	26.7	35100	10	US-09-782-378A-26	Sequence 7680, Ap
c	26	26.8	26.5	1503	10	US-09-815-242-7680	Sequence 61, Appl
c	27	26.8	26.5	5681	9	US-09-974-298-61	Sequence 58, Appl
c	28	26.8	26.5	5681	10	US-09-954-456-721	Sequence 721, App
c	29	26.8	26.5	5994	10	US-09-954-456-1825	Sequence 1825, Ap
c	30	26.8	26.5	5994	10	US-09-954-456-782	Sequence 782, App
c	31	26.6	26.3	6728	10	US-09-954-456-782	Sequence 3946, Ap
c	32	26.6	26.3	6728	10	US-09-880-107-3946	Sequence 15673, A
c	33	26.4	26.1	266	10	US-09-876-574-15673	Sequence 3, Appl
c	34	26.4	26.1	1041	10	US-09-916-790-3	Sequence 1, Appl
c	35	26.4	26.1	1275	10	US-09-734-032-1	Sequence 1, Appl
c	36	26.4	26.1	1275	12	US-10-016-985-1	Sequence 1, Appl
c	37	26.4	26.1	1360	10	US-09-880-107-2269	Sequence 2269, Ap
c	38	26.4	26.1	1473	12	US-10-016-985-3	Sequence 9, Appl
c	39	26.4	26.1	1902	10	US-09-866-582-9	Sequence 1, Appl
c	40	26.4	26.1	2598	10	US-09-816-094-1	Sequence 1, Appl
c	41	26.4	26.1	2893	10	US-09-916-790-1	Sequence 36, Appl
c	42	26.4	26.1	4041	10	US-09-861-289-36	Sequence 38, Appl
c	43	26.2	25.9	788	10	US-09-861-893-38	Sequence 7713, Ap
c	44	26.2	25.9	1929	10	US-09-815-242-7713	Sequence 3, Appl
c	45	26.2	25.9	7301	10	US-09-816-094-3	

ALIGNMENTS

RESULT 1
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925, 301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474

Query Match 30.1%; Score 30.4; DB 10; Length 3209;
Best Local Similarity 67.2%; Pred. No. 0.87;
Matches 43; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 37 TCACAGTGGATGGCGCCGCTACAGAGCGCTGGATGACATCCGATGCGGCGCTAGACCTGC 96
DB 1263 TCACAGTGGATGGCGCCGCTACAGAGCGCTGGATGACATCCGATGCGGCGCTAGACCTGC 1322
QY 97 GCAA 100
DB 1323 GCAA 1326

RESULT 2
US-09-861-289-32
; Sequence 32, Application US/09861289

Patent No. US20020110897A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600,438US1
: CURRENT APPLICATION NUMBER: US/09/861,289
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 32
: LENGTH: 11220
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 29.5%; Score 29.8; DB 10; Length 11220;
Best Local Similarity 66.2%; Pred. No. 1.5;
Matches 43; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 37 TCGACGTGATGCGCCGCTCAGACGCTGGGATGAGTCTCGCTATGGGGCTAGAGCTGC 96
DB 4400 TCGACAGCGGGCGGCTTCCTGCTGACTCGGATTCGACTCGCTGACGGGCTGAGCTCC 4459

QY 97 GCAAC 101
DB 4460 GCAAC 4464

RESULT 3
US-09-861-289-5
: Sequence 5, Application US/09861289
: Patent No. US20020110897A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600,438US1
: CURRENT APPLICATION NUMBER: US/09/861,289
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 36778
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 29.5%; Score 29.8; DB 10; Length 36778;
Best Local Similarity 66.2%; Pred. No. 1.6;
Matches 43; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 37 TCGACGTGATGCGCCGCTCAGACGCTGGGATGAGTCTCGCTATGGGGCTAGAGCTGC 96
DB 20087 TCGACAGCGGGCGGCTTCCTGCTGACTCGGATTCGACTCGCTGACGGGCTGAGCTCC 20146

QY 97 GCAAC 101
DB 20147 GCAAC 20151

RESULT 4
US-09-764-898-58
: Sequence 58, Application US/09764898
: Patent No. US20020090673A1

: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P1201
: CURRENT APPLICATION NUMBER: US/09/764,898
: CURRENT FILING DATE: 2001-01-17
: PRIOR APPLICATION data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 311
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 58
: LENGTH: 1007
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-898-58

Query Match 28.7%; Score 29; DB 10; Length 1007;
Best Local Similarity 57.0%; Pred. No. 2;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 8 GCAGTCTGCGCTCCCGGAGGCAAGCTGAGTGGAGCGCCGCTCAGAGCTGGG 67
DB 590 GCAGTCTGCGCGGCTGCGGCTGCGGAGAGCGGAGCTGAGCGAGCTGGGCGATCGCG 649

QY 68 AATGACTCGCTGATGGGGCTAGAGCTGGCAA 100
DB 650 CGAGAGCGGCTGGCGGCGCGGCGGAGCGCA 682

RESULT 5
US-09-861-289-1
: Sequence 1, Application US/09861289
: Patent No. US20020110897A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600,438US1
: CURRENT APPLICATION NUMBER: US/09/861,289
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 15872
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 28.7%; Score 29; DB 10; Length 15872;
Best Local Similarity 58.8%; Pred. No. 2.6;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 17 GCGCTCCCGGAGCAAGCTGACGTGATGCGCCGCTCAGAGCTGGGAATGAGCTC 76
DB 8233 GCAGCGCCACATCCCGCGGGGCGGAGCGGCTTCCTCAAGACCTGCGCTCGACTC 8292

QY 77 GCTGATGGGCTAGAGCTGGCAAC 101
DB 8293 CTTGGCGGGGTGAGACTCCGCAAC 8317

RESULT 6
US-09-294-0938-5935
: Sequence 5935, Application US/092940938
: Patent No. US2001005135A1
: GENERAL INFORMATION:
: APPLICANT: Laljudi, Raghunath, V.
: APPLICANT: Ito, Laura, Y.
: APPLICANT: Sherman, Bradley, K.
: TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL

```
FILE REFERENCE: PL-0009 US
CURRENT APPLICATION NUMBER: US/09/294,093B
CURRENT FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/082,567
PRIOR FILING DATE: April 21, 1998
NUMBER OF SEQ ID NOS: 6207
SOFTWARE: PERL Program
SEQ ID NO 5935
LENGTH: 290
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20010051335A1 700382685H1
NAME/KEY: unsure
LOCATION: 29
OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-5935
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 290;
Best Local Similarity 58.0%; Pred. No. 2.1;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
OY 9 CAGGTGCTGCGGCTCCCGCAAGCAAGCTCGACGTGATGCGCCGCTCAGAGCTGG3A 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40 CTGAGAGCTGCTGCTCCCTGCTGCGAGACCGCGGAGGCTTCGCTTACCAAGTTCCTTA 99
OY 69 ATGACTCGCTGATGGGCTAGAGCTGC 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 100 AAGAACTGCTGATGCGGAGAGAGCTCC 127
```

RESULT 7

```
US-09-954-456-1955
Sequence 1955, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
```

```
APPLICANT: Young, Paul
```

```
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
```

```
TITLE OF INVENTION: Sets
```

```
FILE REFERENCE: 689290-76
```

```
CURRENT APPLICATION NUMBER: US/09/954, 456
```

```
CURRENT FILING DATE: 2001-09-18
```

```
PRIOR APPLICATION NUMBER: US/60/233, 617
```

```
PRIOR FILING DATE: 2000-09-18
```

```
PRIOR APPLICATION NUMBER: US/60/234, 052
```

```
PRIOR FILING DATE: 2000-09-20
```

```
PRIOR APPLICATION NUMBER: US/60/234, 923
```

```
PRIOR FILING DATE: 2000-09-25
```

```
PRIOR APPLICATION NUMBER: US/60/235, 134
```

```
PRIOR FILING DATE: 2000-09-25
```

```
PRIOR APPLICATION NUMBER: US/60/235, 637
```

```
PRIOR FILING DATE: 2000-09-26
```

```
PRIOR APPLICATION NUMBER: US/60/235, 638
```

```
PRIOR FILING DATE: 2000-09-26
```

```
PRIOR APPLICATION NUMBER: US/60/235, 711
```

```
PRIOR FILING DATE: 2000-09-27
```

```
PRIOR APPLICATION NUMBER: US/60/235, 720
```

```
PRIOR FILING DATE: 2000-09-27
```

```
PRIOR APPLICATION NUMBER: US/60/235, 840
```

```
PRIOR FILING DATE: 2000-09-27
```

```
PRIOR APPLICATION NUMBER: US/60/235, 863
```

```
PRIOR FILING DATE: 2000-09-27
```

```
NUMBER OF SEQ ID NOS: 2276
```

```
SOFTWARE: Perl Program
```

```
SEQ ID NO 1955
```

```
LENGTH: 2638
```

```
TYPE: DNA
```

```
ORGANISM: Homo sapiens
```

```
US-09-954-456-1955
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 2638;
```

```
Best Local Similarity 58.0%; Pred. No. 2.6;
```

```
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
OY 9 CAGGTGCTGCGGCTCCCGCAAGCAAGCTCGACGTGATGCGCCGCTCAGAGCTGGGA 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2114 CTGAGAGCTGCTGCTCCCTGCTGCGAGACCGCGGAGGCTTCGCTTACCAAGTTCCTTA 2173
OY 69 ATGACTCGCTGATGGGCTAGAGCTGC 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2174 AAGAACTGCTGATGCGGAGAGAGCTCC 2201
```

RESULT 8

```
US-09-880-107-2412
Sequence 2412, Application US/09880107
Patent No. US20020142981A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: Horne, Darci T.
```

```
APPLICANT: Vockley, Joseph G.
```

```
APPLICANT: Scherf, Uwe
```

```
TITLE OF INVENTION: Gene Logic, Inc.
```

```
FILE REFERENCE: 44921-5028-WO
```

```
CURRENT APPLICATION NUMBER: US/09/880,107
```

```
CURRENT FILING DATE: 2001-06-14
```

```
PRIOR APPLICATION NUMBER: US 60/211,379
```

```
PRIOR FILING DATE: 2000-06-14
```

```
PRIOR APPLICATION NUMBER: US 60/237,054
```

```
PRIOR FILING DATE: 2000-10-02
```

```
NUMBER OF SEQ ID NOS: 3950
```

```
SOFTWARE: PatentIn Ver. 2.1
```

```
SEQ ID NO 2412
```

```
LENGTH: 2125
```

```
TYPE: DNA
```

```
ORGANISM: Homo sapiens
```

```
FEATURE:
```

```
OTHER INFORMATION: Genbank Accession No. US20020142981A1 M86826
```

```
US-09-880-107-2412
```

```
Query Match          27.7%; Score 28; DB 10; Length 2125;
```

```
Best Local Similarity 55.0%; Pred. No. 4.4;
```

```
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
OY 2 GGTCTGCGAGGCTGCGGCTCCCGCAAGCAAGCTCGACGTGATGCGCGGCTCAGAG 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 926 GGGCTGCGCTGCTGCTGCGGCTGCTCCACACAGCCATCCGAGCTGCGGCGGACCTT 985
OY 62 CCTGGAAATGACCTGCTGATGCGGCTAGAGCTGCGCAAC 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 986 CAAGAGCTGCACTTCTGGAAGAGCTGCGAGCTGCGGCGGAC 1025
```

RESULT 9

```
US-10-044-090-104
Sequence 104, Application US/10044090
Patent No. US20020137081A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: Olga Bandman
```

```
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
```

```
FILE REFERENCE: PA-0028 US
```

```
CURRENT APPLICATION NUMBER: US/10/044,090
```

```
CURRENT FILING DATE: 2002-01-09
```

```
NUMBER OF SEQ ID NOS: 850
```

```
SOFTWARE: PERL Program
```

```
SEQ ID NO 104
```

```
LENGTH: 4149
```

```
TYPE: DNA
```

```
ORGANISM: Homo sapiens
```

```
FEATURE:
```

```
NAME/KEY: misc.feature
```

```
OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.19
```

```
NAME/KEY: unsure
```

```
LOCATION: 1492-1495, 1663, 1669, 4012, 4055
```

```
OTHER INFORMATION: a, t, c, g, or other
```

```
US-10-044-090-104
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101

Sequence: 1 agctctgcagctgctgcgc.....tggggctagagctgcgaac 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6C.COMB.seq.*
7: /cgn2_6/ptodata/2/ina/6D.COMB.seq.*
8: /cgn2_6/ptodata/2/ina/6E.COMB.seq.*
9: /cgn2_6/ptodata/2/ina/6F.COMB.seq.*
10: /cgn2_6/ptodata/2/ina/6G.COMB.seq.*
11: /cgn2_6/ptodata/2/ina/6H.COMB.seq.*
12: /cgn2_6/ptodata/2/ina/6I.COMB.seq.*
13: /cgn2_6/ptodata/2/ina/6J.COMB.seq.*
14: /cgn2_6/ptodata/2/ina/6K.COMB.seq.*
15: /cgn2_6/ptodata/2/ina/6L.COMB.seq.*
16: /cgn2_6/ptodata/2/ina/6M.COMB.seq.*
17: /cgn2_6/ptodata/2/ina/6N.COMB.seq.*
18: /cgn2_6/ptodata/2/ina/6O.COMB.seq.*
19: /cgn2_6/ptodata/2/ina/6P.COMB.seq.*
20: /cgn2_6/ptodata/2/ina/6Q.COMB.seq.*
21: /cgn2_6/ptodata/2/ina/6R.COMB.seq.*
22: /cgn2_6/ptodata/2/ina/6S.COMB.seq.*
23: /cgn2_6/ptodata/2/ina/6T.COMB.seq.*
24: /cgn2_6/ptodata/2/ina/6U.COMB.seq.*
25: /cgn2_6/ptodata/2/ina/6V.COMB.seq.*
26: /cgn2_6/ptodata/2/ina/6W.COMB.seq.*
27: /cgn2_6/ptodata/2/ina/6X.COMB.seq.*
28: /cgn2_6/ptodata/2/ina/6Y.COMB.seq.*
29: /cgn2_6/ptodata/2/ina/6Z.COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	71989	US-09-443-501A-2	Sequence 2, Appli
2	99.4	98.4	68750	US-09-335-409-1	Sequence 1, Appli
3	99.4	98.4	68750	US-09-568-102-1	Sequence 1, Appli
4	99.4	98.4	68750	US-09-567-969-1	Sequence 1, Appli
5	99.4	98.4	68750	US-09-568-480-1	Sequence 1, Appli
6	99.4	98.4	68750	US-09-568-486-1	Sequence 1, Appli
7	99.4	98.4	68750	US-09-568-472-1	Sequence 1, Appli
8	99.4	98.4	68750	US-09-567-899-1	Sequence 1, Appli
9	35.2	34.9	33529	US-09-144-085-3	Sequence 3, Appli
10	32.8	32.5	43280	US-08-804-227C-1	Sequence 1, Appli
11	32.6	32.3	20235	US-07-642-734C-3	Sequence 3, Appli
12	32.6	32.3	20235	US-08-439-009A-3	Sequence 3, Appli
13	31.2	30.9	80161	US-09-036-987A-1	Sequence 1, Appli
14	31.2	30.9	80161	US-09-370-700-1	Sequence 1, Appli
15	30.4	30.1	8460	US-08-469-005A-9	Sequence 9, Appli
16	30.4	30.1	8519	US-09-261-907-1	Sequence 1, Appli
17	30.4	30.1	44377	US-08-804-227C-7	Sequence 7, Appli
18	30.4	30.1	44377	US-08-804-198-1	Sequence 1, Appli
19	29.8	29.5	11220	US-09-105-537-5	Sequence 5, Appli
20	29.8	29.5	36778	US-09-105-537-5	Sequence 5, Appli
21	29.8	29.5	38506	US-09-320-878-19	Sequence 19, Appli
22	29.8	29.5	4403765	US-09-103-840A-2	Sequence 2, Appli
23	29.8	29.5	441529	US-09-103-840A-1	Sequence 1, Appli
24	29.4	29.1	13987	US-08-804-227C-13	Sequence 13, Appli
25	29.2	28.9	1734	US-07-642-734C-1	Sequence 1, Appli
26	29.2	28.7	11219	US-07-642-734C-1	Sequence 1, Appli
27	29	28.7	11219	US-08-439-009A-1	Sequence 1, Appli

28	29	28.7	15872	US-09-105-537-1	Sequence 1, Appli
29	28.4	28.1	328	US-09-060-756-702	Sequence 702, App
30	28.4	28.1	491	US-09-060-756-702	Sequence 107, App
31	28.4	28.1	8501	US-08-793-900-1	Sequence 1, Appli
32	27.6	27.3	245	US-09-404-879A-305	Sequence 305, App
33	27.6	27.1	2588	US-08-796-414B-6	Sequence 6, Appli
34	27.2	26.9	868	US-09-376-728-3	Sequence 3, Appli
35	27.2	26.9	1399	US-08-471-033-24	Sequence 24, Appli
36	27.2	26.9	1399	US-08-471-044-24	Sequence 24, Appli
37	27.2	26.9	1399	US-08-463-483A-24	Sequence 24, Appli
38	27.2	26.9	1399	US-08-471-046A-24	Sequence 24, Appli
39	27.2	26.9	1399	US-08-470-566B-24	Sequence 24, Appli
40	27.2	26.9	1399	US-08-469-334-24	Sequence 24, Appli
41	27.2	26.9	1399	US-09-300-529-24	Sequence 24, Appli
42	27	26.7	35081	US-08-752-760A-1	Sequence 1, Appli
43	27	26.7	50937	US-09-428-517-1	Sequence 1, Appli
44	26.6	26.3	900	PCT-US95-04801-3	Sequence 3, Appli
45	26.6	26.3	3181	US-08-655-086-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kusan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epithilone and Epithilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match      100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCTTCGAGGTGCTGCGCCCTCCCGAAGCAAGCACTGAGCTGAGTCCGCGTCACAGA 60
DB      37570 AGCTTCGAGGTGCTGCGCCCTCCCGAAGCAAGCACTGAGCTGAGTCCGCGTCACAGA 37629
QY      61 GCGTCGAGTGCATCGCTGATGAGGCTAGAGCTCGCAAC 101
DB      37630 GCGTCGAGTGCATCGCTGATGAGGCTAGAGCTCGCAAC 37670

RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409
```

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 43182 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 43241
```

```
OY 61 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 43242 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 43182 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 43241
```

```
OY 61 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 43242 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 43182 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 43241
```

```
OY 61 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 43242 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 43182 AGGCTCGCAGGTGCTGCGCCTCCCGAAGCAAGCTCGACGTGATGCGCGCTCACA 43241
```

```
OY 61 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 43242 GCCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCGCAAC 43282
```



```

RESULT: 6
US-09-568-486-1
: Sequence 1, Application US/09568486
: Patent No. 6355459
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIIONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,486
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-568-486-1

```

	Query Match	Best Local Similarity	99.0%	Score 99.4%	DB 4	Length 68750
	Matches 100%	Conservative	0	Mismatches 1	Indels 0	Gaps 0
OY	1	AGCTTGCAGTGTGCTGGCCCTCCCGAAAGCAACTGACGTGATGCGCCGCTCACA	60			
DB	43182	AGGTTCCAGAGTGTGCTGGCCCTCTCCGAAGGCAACTGACGTGATGCGCCGCTCACA	43241			
OY	61	GCTTGGGAATGACTGCTGATGGGGCTAGACTGCGCAAC	101			
DB	43242	GCTTGGGAATGACTGCTGATGGGGCTAGACTGCGCAAC	43282			

```

RESULT 7
US-09-568-472-1
: Sequence 1, Application US/09568472
: Patent No. 6358719
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zitzke, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-568-472-1

```

		98.4%;	Score 99.4;	DB 4	Length 68750;	
Query Match			Pred. No. 7,9e-19;			
Best Local Similarity						
Matches 100;	Conservative	0;	Mismatches 1;	Indels	0;	Gaps 0;
<hr/>						
OY	1	AGGTCGCCACGACTGCTGGCGCCTCCCGCAAGCACACTGCAGTGGATGGCCGTCTCACA	GA	60		
Db	43182	AGGTCGCCACGACTGCTGGCGCCTCTCCGAAAGCACACTGCAGTGGATGGCCGTCTCACA	GA	43241		
<hr/>						
OY	61	GCCTGCGAATGCACCTGCCTGATTGGGGCTAGAGACTCGCGAAC	CT	101		

Db 43242 GCCTGGGATGGACTCGCTGATGGGGCTAGAGCTGGCCAAC 43282

```

RESULT 8
US-09-567-899-1
: Sequence 1, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567, 899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-899-1

```

	Query Match	98.4%	Score 99.4	DB 4	Length 68750
	Best Local Similarity	99.0%	Pred. No. 7.9e-19		
	Matches 100; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	1	AGGTCGCAGATGGTGGCGCCGCCGAAGAAGCAACTGACGTGGATGGCGCGCTCACA	60		
Db	43182	AGGTCGCAGATGGTGGCGCCCTCTCGAAGGCACACTGCACGTGGATGGCGCGCTCACGA	43241		
OY	61	GCTTGGAATGACCTGCTGATGGGGCTTAGAGCTGGCAAC	101		
Db	43242	GCTTGGAATGACCTGCTGATGGGGCTTAGAGCTGGCAAC	43282		

```

RESULT 9
US-09-144-085-3
: Sequence 3, Application US/09144085
: Patent No. 6280999
: GENERAL INFORMATION:
: APPLICANT: Gustafsson, Claes
: APPLICANT: Betlach, Mary C.
: APPLICANT: Ashley, Gary
: APPLICANT: Julien, Bryan
: APPLICANT: Ziemann, Rainer
: TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 30062-20020. 20
: CURRENT APPLICATION NUMBER: US/09/144,085
: CURRENT FILING DATE: 1998-08-31
: EARLIER APPLICATION NUMBER: 09/010,809
: EARLIER FILING DATE: 1998-01-22
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 33529
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: IS-09-144-085-3

```

	Query Match	34.9%	Score 35.2	DB 4	Length 33529
	Best Local Similarity	65.4%	Pred. No. 0.22		
	Matches 68	Conservative 0	Mismatches 33	Indels 3	Gaps 1
0y	1 AGGTCCTCCAGTGTGCGCTCCCCGAAAGCAAGCTGACGTGGATGCGC---CGCTCA	57			

[illegible]

```

US-08-804-227C-1
US-RESULT 10
Sequence 1 Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kubstoss, Stuart A.
APPLICANT: Rosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

```

Query Match	32.58;	Score 32.8;	DB 2;	Length 43280;
Best Local Similarity	58.08;	Pred. No. 0.99;		
Matches 58;	Conservative 0;	Mismatches 42;	Indels 0;	Gaps 0;

[illegible]

Db 30770 CCTCGGCTTCGACTCGCTCACCGCCGTGAGATGCCAAC 30809

```

RESULT 11
US-07-642-734C-3
Sequence 3, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/APed-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dancigers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"

```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6034..7026
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function="approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function="gene-eryA"
OTHER INFORMATION: /product="orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
```

```
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3
```

```
Query Match      32.3%; Score 32.6; DB 1; Length 20235;
Best Local Similarity 63.3%; Pred. No. 1.1;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY      23 CCCGAAGCAACCTGACGTGATGCCCGCTCAGAGCCCTGGATGAGCTCGCTGAT 82
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      10308 CTCGGCGGACCACTGCCGCAAGCGCTTCAGAGCACTCGTTGACTCGCTGCC 10367
QY      83 GGGGCTAGAGCTGCCGCAAC 101
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      10368 CGCGGTGAGACTGCCGCAAC 10386
```

```
RESULT 12
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Meaulline, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polypeptides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AD6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952-US.D1
TELEPHONE: 847-938-3137
TELECOMMUNICATION INFORMATION:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylCpsynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span

OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylCpsynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14657..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylCpsynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3

Query Match 32.3%; Score 32.6; DB 3; Length 20235;
Best Local Similarity 63.3%; Pred. No. 1.1;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 : Search time 157 Seconds
(without alignments)
1448.738 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101
Sequence: 1 aggtctgcagtgctgcgc.....tggggctagagctgcgcac 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	71989	21	AAA29349
2	99.4	98.4	68750	21	AAZ55887
3	40.2	39.8	5292	22	AAF90036
4	40.2	39.8	34071	22	AAF90033
5	40.2	39.8	42717	22	AAF90032
6	35.2	34.9	33529	23	AAAF17367
7	34.2	33.9	4615	22	AAF90034
8	33	32.7	6564	22	AAF31637
9	32.8	32.5	43280	18	AA780413

10	32.6	32.3	29879	14	AAO46806	eryA region of S.
11	32.2	31.9	53789	19	AAV21187	Amycolatopsis medi
12	31.6	31.3	77536	21	AAA14651	Nucleotide sequenc
13	31.4	31.1	4466	21	AAA14663	Nucleotide sequenc
14	31.4	31.1	4478	21	AAA14661	Nucleotide sequenc
15	31.4	31.1	4547	21	AAA14664	Nucleotide sequenc
16	31.4	31.1	4571	21	AAA14662	Nucleotide sequenc
17	31.4	31.1	77536	21	AAA14651	Nucleotide sequenc
18	31.2	30.9	6459	22	AAF88336	S. spinosa DNA fra
19	31.2	30.9	7788	22	AAF88335	S. spinosa DNA fra
20	31.2	30.9	50000	22	AAF88312	S. spinosa DNA fra
21	31.2	30.9	50000	22	AAF88313	S. spinosa DNA fra
22	31.2	30.9	50000	22	AAF88316	S. spinosa DNA fra
23	31.2	30.9	80161	20	AAZ21501	DNA fragment of Sa
24	31.2	30.9	125401	22	AAAD17186	Streptomyces nous
25	31	30.7	47981	22	AAF30757	Micromonospora meg
26	30.8	30.5	14806	22	AAAD03809	Streptomyces galii
27	30.6	30.3	1419	21	AAAC55827	S. lavendulae Mmcm
28	30.6	30.3	6462	22	AAF90037	Nucleotide sequenc
29	30.6	30.3	53500	21	AAAC55842	Complete nucleotid
30	30.4	30.1	2237	24	ABL62699	Colon adenocarcino
31	30.4	30.1	3209	21	AACT78080	Human cancer assoc
32	30.4	30.1	5676	19	AAV21186	Amycolatopsis medi
33	30.4	30.1	8460	18	AA788206	cDNA for protein (
34	30.4	30.1	8470	21	AAZ95007	Cancer specific ge
35	30.4	30.1	8519	21	AAA37760	Human fatty acid s
36	30.4	30.1	44377	18	AA78508	Platenolide synth
37	30.4	30.1	44377	18	AA780414	Platenolide synth
38	29.8	29.5	11220	21	AAZ87298	S. venezuelae macr
39	29.8	29.5	14071	23	ABL50560	Micromonospora car
40	29.8	29.5	36778	21	AAZ87318	S. venezuelae pik
41	29.8	29.5	37948	21	AAZ87285	Nucleotide sequenc
42	29.8	29.5	38506	21	AAAF5633	Recombinant cosmid
43	29.8	29.5	38506	21	AAZ56001	Mycobacterium tube
44	29.8	29.5	4403765	22	AA199663	Mycobacterium tube
45	29.8	29.5	4411529	22	AA199682	Mycobacterium tube

ALIGNMENTS

RESULT 1	AAA29349	standard; DNA; 71989 BP.
AAA29349;		
12-SEP-2000 (first entry)		
Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.		
Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;		
epoL; epoK; P450 epoxidase; ORFA; ORB; promoter; enhancer; anti-fungal;		
tubulin polymerization assay; anti-tumour; cytostatic; ds.		
Sorangium cellulosum.		
Location/Qualifiers		
3..992		
key		
CDS		
FT	/*tag= a	
FT	/label= ORF_A	
FT	/product= transposase	
FT	/note= "not part of the PKS"	
FT	989..1501	
CDS		
FT	/*tag= b	
FT	/label= ORF_B	
FT	/product= transposase	
FT	/note= "not part of the PKS"	
FT	1998..6263	
CDS		
FT	/*tag= c	
FT	/label= epoA_gene	
FT	/note= "encodes the loading domain"	
FT	2031..3548	
FT	misc_RNA	

```

FT      /tag- d
FT      /note- "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /tag- e
FT      /note- "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /tag- f
FT      /note- "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /tag- g
FT      /note- "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /tag- h
FT      /label- epob_gene
FT      /note- "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /tag- i
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /tag- j
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /tag- k
FT      /note- "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /tag- l
FT      /note- "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /tag- m
FT      /note- "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag- n
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag- o
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /tag- p
FT      /note- "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /tag- q
FT      /note- "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /tag- r
FT      /note- "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /tag- s
FT      /note- "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /tag- t
FT      /note- "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /tag- u
FT      /note- "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /tag- v
FT      /note- "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /tag- w
FT      /note- "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /tag- x

```

```

FT      /label- epoc_gene
FT      /note- "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /tag- y
FT      /note- "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /tag- z
FT      /note- "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /tag- aa
FT      /note- "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /tag- ab
FT      /note- "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /tag- ac
FT      /note- "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /tag- ad
FT      /label- epod_gene
FT      /note- "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /tag- ae
FT      /note- "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /tag- af
FT      /note- "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /tag- ag
FT      /note- "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /tag- ah
FT      /note- "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /tag- ai
FT      /note- "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /tag- aj
FT      /note- "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /tag- ak
FT      /note- "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /tag- al
FT      /note- "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /tag- am
FT      /note- "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /tag- an
FT      /note- "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /tag- ao
FT      /note- "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /tag- ap
FT      /note- "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /tag- aq
FT      /note- "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /tag- ar
FT      /note- "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /tag- as
FT      /note- "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /tag- at
FT      /note- "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /tag- au

```



```

FT      /note= "encodes DH6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /tag= av
FT      /note= "encodes KR6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= epof_gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      41341..41922
FT      /tag= db
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match          100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.2e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGGTCGCGAGGTGCTGGCCCTCCCGAAGCAAGCTGAGCTGGATCGCGCTCAGCA 60
Db      37570 AGGTCGCGAGGTGCTGGCCCTCCCGAAGCAAGCTGAGCTGGATCGCGCTCAGCA 37629
QY      61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGCAAC 101
Db      37630 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGCAAC 37670

RESULT 2
ID      AA255887 standard; DNA; 68750 BP.
XX      AA255887;
XX      10-APR-2000 (first entry)
DE      Sorangium cellulosum 68.75 kb contig.
XX      Epithelone biosynthesis; type I polyketide synthase; taxol substitute;
KW      anticancer; ds.
XX      Sorangium cellulosum.
OS      Key
FH      Location/Qualifiers
FT      1..1826
FT      /tag= a
FT      /partial
FT      /product= "Partial Orf 1 protein (AAV58580)"
FT      /note= "No initiation codon given in the specification"
FT      complement (1900..3171)
FT      /tag= b
FT      /product= "Orf 2 protein (AAV58581)"
FT      3415..5556
FT      /tag= c
FT      /product= "Orf 3 protein (AAV58582)"
FT      complement (5612..5992)
FT      /tag= d
FT      /product= "Orf 4 protein (AAV58583)"
FT      6226..6675
FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
FT      /tag= f

```

```

FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /partial
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"

MO9966028-A2.
PD      23-DEC-1999.
XX      16-JUN-1999; 99WO-EP04171.
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
PR      05-FEB-1999; 99US-0118906.
XX      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX      Schnupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;

```

```

XX WPI: 2000-097741/08.
DR P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.
XX
XX New isolated epothilone synthase genes, used for the recombinant
XX production of epothilone for use in cancer therapy.
XX
XX Claim 14; Page 87-104; 174pp; English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX comprising 22 open reading frames (ORFs) and includes genes encoding
XX proteins involved in the biosynthesis of epothilones. Epothilones A and
XX B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX starter unit; polyketides being synthesised from two-carbon building
XX blocks, the beta-carbon of which always carries a keto group. Each round
XX of two-carbon addition is carried out by a complex of enzymes known as
XX the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
XX the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
XX and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
XX formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
XX the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
XX involved in transport. Epothilones mimic the biological activity of
XX taxol, and may be substituted for taxol in cancer chemotherapeutic
XX compositions. Epothilones exhibit a much lower drop in potency against a
XX multiply drug-resistant cell line compared with taxol, and are
XX considerably less efficiently exported from such cells by the multidrug
XX resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX epothilones as anticancer agents, they are problematical to produce on a
XX large scale. Epothilones are too complex for industrial scale chemical
XX synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX poor yields of epothilones. The nucleic acids of the invention may be
XX used for the recombinant production of epothilones in a heterologous host
XX that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
SQ
XX
XX Query Match 98.4%; Score 99.4; DB 21; Length 68750;
XX Best Local Similarity 99.0%; Pred. No. 3.5e-19;
XX Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 AGGTCGCGAGGTGTCGCCCTCCCGAAGGCAAGCTCGAGTGGCGCGCTACGCA 60
XX |||||||
XX DB 43182 AGGTCGCGAGGTGTCGCCCTCTCGAAGGCAAGCTCGAGTGGCGCGCTACGCA 43241
XX |||||||
XX QY 61 GCCTGGGAATGACCTGCTGATGGGCGCTAGAGCTGGCGCAAC 101
XX |||||||
XX DB 43242 GCCTGGGAATGACCTGCTGATGGGCGCTAGAGCTGGCGCAAC 43282
XX |||||||
XX
XX RESULT 3
XX AAF90036
XX ID AAF90036 standard; DNA; 5292 BP.
XX
XX AC AAF90036;
XX
XX DT 06-AUG-2001 (first entry)
XX
XX DE Nucleotide sequence of a type I polyketide synthase.
XX
XX KM Metabolic pathway operon; polyketide; polyketide antibiotic;
XX type I polyketide synthase; ss.
XX
XX OS Unidentified.
XX
XX Key Location/Qualifiers
XX FH 1..5292
XX FT /tag= a
XX FT /product= "type I polyketide synthase"
XX

```

```

PN WO200140497-A2.
XX
XX PD 07-JUN-2001.
XX
XX PF 27-NOV-2000; 2000WO-FR03311.
XX
XX PR 29-NOV-1999; 99FR-0015032.
XX PR 07-JUN-2000; 2000US-0209800.
XX
XX PA (AVER ) AVENTIS PHARMA SA.
XX
XX PI Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
XX PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
XX PI Frostegard A;
XX
XX DR WPI: 2001-374849/39.
XX P-PSDB; AAB83973.
XX
XX PT Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics.
XX
XX PS Claim 35; Page 305-307; 356pp; French.
XX
XX CC The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX CC milling a dried sample to produce microparticles; suspending these in
XX CC liquid buffer; extraction of nucleic acids from the microparticle;
XX CC passing nucleic acid-containing solution through a molecular sieve;
XX CC passing nucleic acid-enriched fractions through an anion exchange
XX CC chromatography material; and recovering fractions containing purified
XX CC nucleic acids. The nucleic acids are sources for sequences that encode
XX CC either operons involved in a metabolic pathway (specifically polyketide
XX CC synthesis) or polyketides, particularly for production of therapeutic
XX CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
XX CC represent open reading frames (ORFs) of the coding strand of cosmid
XX CC a2691, and encode type I polyketide synthases.
XX
XX Sequence 5292 BP; 905 A; 1715 C; 1725 G; 947 T; 0 other;
SQ
XX
XX Query Match 39.8%; Score 40.2; DB 22; Length 5292;
XX Best Local Similarity 64.5%; Pred. No. 0.022;
XX Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
XX 9 CAGGTCTGCGCTGCTCCCGAAGGCAAGCTCGAGTGGCGCGCTACGAGCTGGGA 68
XX |||||||
XX DB 4885 CAGGTCTGCGCTGCTGATCGACGGAACGACGAGGCGAGCGGTGATCTCGGA 4944
XX |||||||
XX QY 69 ATGGAATCGCTGATGGGCGCTAGAGCTGGCGCAAC 101
XX |||||||
XX DB 4945 TTGCAATTCGTTGATGGCGCTCGAGTTTCGCAAC 4977
XX |||||||
XX
XX RESULT 4
XX AAF90033
XX ID AAF90033 standard; DNA; 34071 BP.
XX
XX AC AAF90033;
XX
XX DT 06-AUG-2001 (first entry)
XX
XX DE Nucleotide sequence of cosmid a2691 (coding strand).
XX
XX KM Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX OS Synthetic.
XX
XX PN WO200140497-A2.
XX
XX PD 07-JUN-2001.
XX
XX PF 27-NOV-2000; 2000WO-FR03311.
XX

```

PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVER) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
XX
PR Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Example 14; Page 289-300; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a2691 coding strand), which encodes different
CC polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
XX
Query Match 39.8%; Score 40.2; DB 22; Length 34071;
Best Local Similarity 64.5%; Pred. No. 0.026;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
QY 9 CAGGTCGTGCGCCGCCGAGGCAAGCTCGACGTGGCGCTCAGCAGCCTGGGA 68
DB 17820 CACGTGCTGCGCTTGTGATCCAGCGAAGCTCGACGAGCGAGCTGGTATCTCGGA 17879
XX
QY 69 ATGGAAGCTGCTGATGGGCTAGAGCTGCGCAAC 101
DB 17880 TTCGATTCGTGATGGCCCTCGAGTTTCGCAAC 17912
XX
RESULT 5
AAF90032/C
ID AAF90032 standard; DNA; 42717 BP.
XX
AC AAF90032;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of cosmid a2691 (non-coding strand).
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS Synthetic.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVER) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;

PI Frostegard A;
XX
DR WPI: 2001-374849/39.
XX
PR Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Example 14; Page 274-288; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a2691 (non-coding strand). The sense strand
CC encodes different polyketide synthases.
XX
SQ Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
XX
Query Match 39.8%; Score 40.2; DB 22; Length 42717;
Best Local Similarity 64.5%; Pred. No. 0.027;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
QY 9 CAGGTCGTGCGCCGCCGAGGCAAGCTCGACGTGGCGCTCAGCAGCCTGGGA 68
DB 16300 CACGTGCTGCGCTTGTGATCCAGCGAAGCTCGACGAGCGAGCTGGTATCTCGGA 16241
XX
QY 69 ATGGAAGCTGCTGATGGGCTAGAGCTGCGCAAC 101
DB 16240 TTCGATTCGTGATGGCCCTCGAGTTTCGCAAC 16208
XX
RESULT 6
AAS17367
ID AAS17367 standard; DNA; 33529 BP.
XX
AC AAS17367;
XX
DT 12-MAR-2002 (first entry)
XX
DE DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW acyl transferase domain; dehydratase domain; ketoreductase domain;
KW acyl carrier protein domain; pKOS28-26; ds.
XX
OS Sorangium cellulosum.
XX
PN US6280999-B1.
XX
PD 28-AUG-2001.
XX
PF 31-AUG-1998; 98US-0144085.
XX
PR 22-JAN-1998; 98US-0010809.
XX
PA (KOSA-) KOSAN BIOSCIENCE.
XX
PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX
DR WPI: 2001-606536/69.
XX
PR Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT polyketide open reading frame encoding modules with one or more domains
PT such as ketosynthase, acyl transferase and acyl carrier protein domains

```
XX Claim 4: Fig 1: 72pp; English.
PS
XX
CC The present invention relates to the isolation of novel Sorangium
CC cellulose polyketide synthases (PKS), and the polynucleotide sequences
CC encoding them. The polyketide synthases include catalytic domains such
CC as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC ketoreductase domain and acyl carrier protein domain. A host cell
CC comprising a PKS ORF (open reading frame) which encodes one or more
CC more PKS domains is useful for producing polyketide synthases from which
CC polyketides can be produced. The host cells are useful for constructing
CC a library, where each individual colony of the library represents a
CC colony with the ability to produce a particular PKS synthase and
CC ultimately a particular polyketide. The polyketides produced by these
CC colonies can be used collectively in a panel to represent a library or
CC may be assessed individually for activity. Colonies in the library are
CC also induced to produce the relevant synthases and thus to produce the
CC relevant polyketides to obtain a library of candidate polyketides which
CC can be screened for binding to desired targets such as receptors,
CC signalling proteins, etc. The present sequence represents the DNA
CC sequence of cosmid pK0528-26 which encodes one or more domains of
CC S. celluloseum PKS.
CC Note: The present sequence is said to encode the functional domains
CC of S. celluloseum PKS which correspond to domains or domain subsets of
CC the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other:

Query Match      34.9%; Score 35.2; DB 23; Length 33529;
Best Local Similarity 65.4%; Pred. No. 0.7;
Matches 68; Conservatve 0; Mismatches 33; Indels 3; Gaps 1;

QY 1 AGGTCGCGAGTGTCTGCGCCCTCCCGAAGCAAGCTGACGTGATGCGC---CGCTCA 57
   ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
DB 4830 AGTCCGCCGCGGTCTCGGCGCTCCAGCGACGAGCGCCCGTGGGGGCGAGCAGTCTCTGA 4889
   | |||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
QY 58 CGAGCTGGGAATGAGACTGCTGATGGGCTAAGAGCTGCCCAAC 101
   | |||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
DB 4890 AGGGCTCTGGGCTGACCTGATGGCGGTGGAGCTGCGCAAC 4933

RESULT 7
AAF90034
ID AAF90034 standard; DNA: 4615 BP.
XX
AC AAF90034;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
KW Metabolic pathway operon: polyketide; polyketide antibiotic;
KM type I polyketide synthase; ss.
XX
OS unidentified.
XX
FH Key Location/Qualifiers
FT CDS 2..4615
FT /tag= a
FT /product= "type I polyketide synthase"
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR03311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVER ) AVENTIS PHARMA SA.
XX
XX Jeannin P, Pernodet J, Guerin M, Simonet P, Courtois S,
PI
```

```
PI Cappellano C, Francou F, Raynal A, Bail M, Seznov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
DR P-PSDB: AAB63971.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Claim 35; Page 300-302; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a2691, and encode type I polyketide synthases.
XX
SQ Sequence 4615 BP; 827 A; 1477 C; 1429 G; 882 T; 0 other:

Query Match      33.9%; Score 34.2; DB 22; Length 4615;
Best Local Similarity 62.1%; Pred. No. 1.1;
Matches 54; Conservatve 0; Mismatches 33; Indels 0; Gaps 0;

QY 15 CTGCGCTCCCGAAGCAAGCTGACGTGATGCGCGCTCAGCAGCCTGGGAATGAC 74
   ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
DB 2 CTGAGTGGCCCGAATGCGGCGGTGAGCTTACAGAGCGCGTGGGTATGGATTGGAC 61
   | |||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
QY 75 TCGCTATGGGCGCTAGAGCTGCCCAAC 101
   | |||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
DB 62 TCGCTATGGGCGGTGCAATTACGCCAAC 88

RESULT 8
AAF31637
ID AAF31637 standard; DNA: 6564 BP.
XX
AC AAF31637;
XX
DT 09-APR-2001 (first entry)
XX
DE Mycobacterium tuberculosis ppsc gene.
XX
XX Mycobacterium tuberculosis; attenuated microorganism;
KM signature tagged transposon mutant; mutant library;
KM mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW vaccine; ppsc; ds.
XX
XX Mycobacterium tuberculosis.
XX
OS
XX
XX WO200102555-A1.
XX
XX 11-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-IB00950.
XX
XX 06-JUL-1999; 99US-0142982.
XX 08-JUL-1999; 99US-0142833.
XX
XX (INSP ) INSP PASTEUR.
XX
XX Gicquel B, Guilhot C, Camacho L;
XX
XX WPI: 2001-091804/10.
XX P-PSDB: AAB66467.
DR
```

XX Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
PT comprises using signature tagged transposon mutagenesis -
XX
XX Example 8; Page 122-124; 159pp; English.
XX The present sequence is given in a specification relating to a method for
XX screening a library of mutants. The method comprises constructing a
XX library with insertions in genes and/or regulatory regions of the
XX organisms of interest, where the insertion contains a tag and/or a
XX transposon associated with a tag. The mutants are identified by
XX hybridisation of the tags to known sequences. The method is useful for
XX treating an individual suffering from a mycobacterial infection,
XX suspected of being infected with a Mycobacterium, or having been
XX exposed to an infectious Mycobacterium. It is also useful for
XX identifying and isolating mutants of actinomycetales and for identifying
XX compounds that have antibiotic activity. The method is used to identify
XX mutants of microorganisms, preferably an actinomycetales, such as
XX M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
XX M. paratuberculosis, that is unable to grow under specific conditions.
XX It is especially useful for identifying loci involved in pathogenicity.
XX It is useful in constructing vaccines. The method can be used to screen
XX multiple libraries concurrently. It can screen libraries of different
XX organisms or different strains of the same organism.
XX
XX Sequence 6564 BP; 1137 A; 2189 C; 2154 G; 1084 T; 0 other;
SQ
Query Match 32.7%; Score 33; DB 22; Length 6564;
Best Local Similarity 60.7%; Pred. NO. 2.6;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
OY 12 GTGTCGCCCTCCCGGAGCGACGCTGATGCGCCCTCAGCGCTGGGATG 71
DB 6250 GCGGTGCTGCTGCGGCGATCCATCGATCCACCGACCGCTGGAACCTGGGACTTC 6309
OY 72 GACTCGCTGATGGGCTAGAGCTGGCAA 100
DB 6310 GACTCGCTGATGGGCTGGAAATGGCGAA 6338
RESULT 9
AAT80413
ID AAT80413 standard; DNA: 43280 BP.
XX
XX AAT80413;
XX
XX 27-FEB-1998 (first entry)
XX
XX DE Tylactone synthase gene cluster.
XX
XX KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
XX polypeptide; tylactone synthesis; antibiotic; tylosin; ss.
XX
XX OS Streptomyces fradiae.
XX
XX
XX Key location/Qualifiers
FH 816..14243
FT /*tag= a
FT /transl_except= (pos: 816..818, aa: Met)
FT /note= "ORF1 encodes protein shown in AAW22601"
FT 14351..19945
FT /*tag= b
FT /transl_except= (pos: 14351..14353, aa: Met)
FT /note= "ORF2 encodes protein shown in AAW22602"
FT 20010..31199
FT /*tag= c
FT /transl_except= (pos: 20010..20012, aa: Met)
FT /note= "ORF3 encodes protein shown in AAW22603"
FT 31232..36067
FT /*tag= d
FT /note= "ORF4 encodes protein shown in AAW22604"
FT 36249..41774
FT CDS

FT /*tag= e
FT /note= "ORF5 encodes protein shown in AAW22605"
XX
XX EP791655-A2.
XX
XX PD 27-AUG-1997.
XX
XX PF 19-FEB-1997; 97EP-0301056.
XX
XX PR 22-FEB-1996; 96US-0012078.
XX
XX PA (ELIL) LILLY & CO ELI.
XX
XX PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
XX
XX DR WPI: 1997-418046/39.
XX
XX DR P-PSDB: AAW22601-W22605.
XX
XX PT DNA encoding Streptomyces fradiae tylosin synthase domain - for
XX production of tylosin-related polyketide compounds
XX
XX PS Claim 2; Pages 8-66; 220pp; English.
XX
XX CC This sequence represents the tylosin synthase gene cluster of the
XX invention. This sequence is also referred to as the tylG gene, and was
XX isolated from Streptomyces fradiae. This sequence encodes multifunctional
XX proteins which direct the synthesis of the polyketide tylosin, isolated
XX from Streptomyces fradiae. Tylosin is the basic building block of the
XX antibiotic tylosin. The DNA sequence can be modified so as to alter the
XX type of carboxylic acids incorporated, the number of carboxylic acids
XX incorporated and/or the post-condensation reactions performed, thereby
XX resulting in novel tylosin-related polyketides.
XX
XX SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
SQ
Query Match 32.5%; Score 32.8; DB 18; Length 43280;
Best Local Similarity 58.0%; Pred. NO. 3.5;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
OY 2 GGTCTGCGAGTGTGCGCTCCCGAGGCAAGCTCGAGCTGATGGCGCTCAGCAG 61
DB 30710 GCGGCGGTGCTGGGCGGCGATCCGACGATCGATCCCGGAGCGGCTTCAGGA 30769
OY 62 CCTGGGAATGAGCTGCTGATGGGCTAGAGCTGGCAAC 101
DB 30770 CCTGGGCTTGACTGCTCACCCTGGGATGGGCAAC 30809
RESULT 10
AAQ46806
ID AAQ46806 standard; DNA: 29879 BP.
XX
XX AC AAQ46806;
XX
XX XX 22-DEC-1993 (first entry)
XX
XX DE eryA region of S. erythraea chromosome.
XX
XX KW Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
XX erythromycin; condensation; elongation; acyl chain growth;
XX gene replacement; ss.
XX
XX OS Saccharopolyspora erythraea.
XX
XX
XX Key location/Qualifiers
FH 3..10199
FT /*tag= a
FT /note= "ORF 1"
FT 10218..20921
FT /*tag= b
FT /note= "ORF 2"
FT 20922..29879
FT /*tag= c
FT CDS

```

FT XX /note= "ORF 3"
FT XX WO9313663-A.
XX
XX 22-JUL-1993.
PD
XX 17-JAN-1992: 92MO-US00427.
XX
XX 17-JAN-1992: 92MO-US00427.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Donadio S, Katz L, McAlpine JB;
XX
XX MPI: 1993-242804/30.
XX P-PSDB: AAR44430-32.
XX
XX Biosynthesis of specific polyketide analogues esp. erythromycin
XX cpds. - by introducing altered biosynthetic gene-contg. DNA into
XX microorganisms
XX
XX Claim 27; Fig 2; 13pp; English.
XX
XX This sequence represents a fragment of the Saccharopolyspora erythraea
XX genome, designated eryA. The polypeptides encoded by this region
XX are involved in the biosynthesis of the polyketide segment of
XX erythromycin. eryA is organised in modules and each module takes care
XX of one condensation step. The precise succession of elongation steps
XX is dictated by the genetic order of the modules. This fragment may be
XX specifically altered such that novel polyketide molecules of desired
XX structure are produced. Three types of alteration may be produced;
XX those inactivating a single function in a module which does not arrest
XX acyl chain growth; those inactivating a single function in a module
XX which does affect chain growth; and those affecting an entire module.
XX The mutations may be introduced by gene replacement.
XX
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
SQ
Query Match 32.3%; Score 32.6; DB 14; Length 29879;
Best Local Similarity 63.3%; Pred. No. 3.9;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY 23 CCCCCAAGGCAAGCTGACGTGGATGCCGCTCAGAGCTTGGGAATGATCGCTGAT 82
DB 20507 CTCGGCCGACACCTCCCGCAGGCAAGCGCTTCAAGACCTCGCTTGCATCGCTGGC 20566
OY 83 GGGGCTAGAGCTCGCGCAAC 101
DB 20567 CGCGGTGAGCTCGCGCAAC 20585
RESULT 11
AAV21187
ID AAV21187 standard; DNA: 53789 BP.
XX
XX AAV21187;
XX
XX 24-JUL-1998 (first entry)
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
XX Amycolatopsis mediterranei rifamycin synthesis; gene cluster;
XX KW polyketide synthase; actinomycete; ansamycin; ds.
XX
XX Amycolatopsis mediterranei.
XX
XX Key Location/Qualifiers
XX CDS 1825..15543
XX FT /tag= a
XX FT /label= ORF_A
XX FT /product= "polyketide synthase"
XX CDS 15550..30759
XX FT /tag= b

```

```

FT FT /label= ORF_B
FT FT /product= "polyketide synthase"
FT FT 30895..36060
FT CDS /tag= c
FT FT /label= ORF_C
FT FT /product= "polyketide synthase"
FT FT 36259..41325
FT CDS /tag= d
FT FT /label= ORF_D
FT FT /product= "polyketide synthase"
FT FT 41373..51614
FT FT /tag= e
FT FT /label= ORF_E
FT FT /product= "polyketide synthase"
FT FT 51713..5293
FT CDS /tag= f
FT FT /label= ORF_F
FT FT /product= "polyketide synthase"
XX
XX WO9807868-A1.
XX
XX 26-FEB-1998.
XX
XX 18-AUG-1997: 97MO-EP04495.
XX
XX 20-AUG-1996: 96EP-0810551.
XX
XX (NOVS ) NOVARTIS AG.
XX
XX Engel N, Schupp T, Toupet C;
XX
XX MPI: 1998-169172/15.
XX P-PSDB: AAM52845-W52850.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX to produce rifamycin and rifamycin analogues
XX
XX Claim 4; Page 53-102; 205pp; English.
XX
XX The present sequence represents a Amycolatopsis mediterranei rifamycin
XX synthesis gene cluster DNA fragment from the present invention. The
XX DNA fragment comprises a DNA region involved directly or indirectly
XX in the gene cluster responsible for rifamycin synthesis, including
XX the adjacent DNA regions to the right and left which, by reason of
XX their function in connection with rifamycin biosynthesis, qualify
XX as constituents of this rifamycin gene cluster, and functional
XX fragments, derivatives or constituents of these. The Amycolatopsis
XX mediterranei rifamycin synthesis gene cluster DNA fragment can be used
XX for producing rifamycin, rifamycin analogues or precursors. It can also
XX be used for inactivating or modifying genes involved in ansamycin or
XX rifamycin biosynthesis. The DNA can be used for constructing mutant
XX actinomycetes strains from which the natural rifamycin or ansamycin
XX biosynthesis gene cluster has been partly or completely deleted. The
XX DNA fragment can be used for assembling a library of polyketide
XX synthases, which can be used for assembling a library of polyketides.
XX A hybridisation probe of the invention can be used for identifying DNA
XX fragments involved in the biosynthesis of ansamycins.
XX
XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;
SQ
Query Match 31.9%; Score 32.2; DB 19; Length 53789;
Best Local Similarity 61.2%; Pred. No. 5.3;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY 17 GCGCCTCCCGGAGGCAAGCTGACGTGGATGCCGCTCAGAGCTTGGGAATGACATC 76
DB 10617 GCACCGGGAGCGCGGAGTTTCAGACTCGACGCGCTTCAAGACCGGCTTGCATC 10676
OY 77 GCTGATGGGCGCTAGAGCTCGCAAC 101
DB 10677 GCTACGCGGCGTCAAGCTCGCGCAAC 10701

```

```
RESULT 12
AAAA14651
ID AAA14651 standard; DNA: 77536 BP.
XX
XX AAA14651;
XX
XX 08-AUG-2000 (first entry)
XX
XX Nucleotide sequence of the FK-520 biosynthetic gene cluster.
DE
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW FK-506 binding protein; polyketide compound; transplant rejection;
KW graft-versus-host disease; uveitis; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
XX Streptomyces hygroscopicus.
XX
XX OS
XX FH Key
XX FT Location/Qualifiers
XX FT complement (412..1836)
XX FT /tag= a
XX FT /note= "fkpW gene"
XX FT complement (2020..3579)
XX FT /tag= b
XX FT /note= "fkpV gene"
XX FT 3969..4496
XX FT /tag= c
XX FT /note= "fkpR2 gene"
XX FT complement (4595..5488)
XX FT /tag= d
XX FT /note= "fkpR1 gene"
XX FT 5601..6818
XX FT /tag= e
XX FT /note= "fkpE gene"
XX FT 6808..8052
XX FT /tag= f
XX FT /note= "fkpF gene"
XX FT 8156..8824
XX FT /tag= g
XX FT /note= "fkpG gene"
XX FT complement (9122..9883)
XX FT /tag= h
XX FT /note= "fkpH gene"
XX FT complement (9894..10994)
XX FT /tag= i
XX FT /note= "fkpI gene"
XX FT complement (10987..11247)
XX FT /tag= j
XX FT /note= "fkpJ gene"
XX FT complement (11244..12092)
XX FT /tag= k
XX FT /note= "fkpK gene"
XX FT complement (12113..13150)
XX FT /tag= l
XX FT /note= "fkpL gene"
XX FT complement (13212..23988)
XX FT /tag= m
XX FT /note= "fkpC gene"
XX FT complement (13452..13662)
XX FT /tag= n
XX FT /note= "ACP6"
XX FT complement (13761..14394)
XX FT /tag= o
XX FT /note= "KR6"
XX FT complement (14517..15294)
XX FT /tag= p
XX FT /note= "ER6"
XX FT complement (15438..16587)
XX FT /tag= q
XX FT
XX FT misc_feature
XX FT /note= "dehydratase domain (DH) 6"
XX FT complement (16587..17820)
XX FT /tag= r
XX FT /note= "acyltransferase domain (AT) 6"
XX FT complement (17820..19053)
XX FT /tag= s
XX FT /note= "KS6"
XX FT complement (19116..19326)
XX FT /tag= t
XX FT /note= "ACP5"
XX FT complement (19464..20097)
XX FT /tag= u
XX FT /note= "KR5"
XX FT complement (20241..21420)
XX FT /tag= v
XX FT /note= "DH5"
XX FT complement (21420..22653)
XX FT /tag= w
XX FT /note= "AT5"
XX FT complement (22653..23892)
XX FT /tag= x
XX FT /note= "KS5"
XX FT complement (23992..46573)
XX FT /tag= y
XX FT /note= "fkpB gene"
XX FT complement (24163..24373)
XX FT /tag= z
XX FT /note= "ACP4"
XX FT complement (24997..26146)
XX FT /tag= aa
XX FT /note= "DH4 (inactive)"
XX FT complement (26146..27430)
XX FT /tag= ab
XX FT /note= "AT4"
XX FT complement (27430..28684)
XX FT /tag= ac
XX FT /note= "KS4"
XX FT complement (28750..28960)
XX FT /tag= ad
XX FT /note= "ACP3"
XX FT complement (29092..29740)
XX FT /tag= ae
XX FT /note= "KR3"
XX FT complement (29869..31018)
XX FT /tag= af
XX FT /note= "DH3 (inactive)"
XX FT complement (31018..32185)
XX FT /tag= ag
XX FT /note= "AT3"
XX FT complement (32185..33439)
XX FT /tag= ah
XX FT /note= "KS3"
XX FT complement (33505..33715)
XX FT /tag= ai
XX FT /note= "ACP2"
XX FT complement (33823..34480)
XX FT /tag= aj
XX FT /note= "KR2"
XX FT complement (34606..35749)
XX FT /tag= ak
XX FT complement (28750..28960)
XX FT /tag= al
XX FT /note= "DH2 (inactive)"
XX FT complement (35749..37144)
XX FT /tag= am
XX FT /note= "AT2"
XX FT complement (37145..38296)
XX FT /tag= an
XX FT /note= "KS2"
XX FT complement (38371..38581)
XX FT /tag= ao
XX FT /note= "ACP1"
XX FT complement (38677..39307)
XX FT
XX FT misc_feature
```

[illegible]

RESULT 13
AAA14663
ID AAA14663 standard; DNA; 4466 BP

XX	AAAI4663;
AC	
XX	
DT	08-AUG-2000 (first entry)
XX	
DE	Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX	
KW	FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW	transnaphthyl; FK-506 binding protein; polyketide compound; uveitis;
KW	immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW	transplant rejection; graft-versus-host disease; alopecia universalis;
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;
KW	multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW	neutrite outgrowth; nerve regrowth; Parkinson's disease;
KW	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KM	peripheral neuropathy; ss.
XX	
OS	Synthetic.
OS	Streptomyces hygroscopicus.
XX	
FH	Key
XX	
FT	location/Qualifiers
FT	9..4434
FT	/*tag= a
XX	/note= "no termination codon given"
XX	
PN	MO200020601-A2.
XX	
PD	13-APR-2000.
XX	
PF	01-OCT-1999; 99WO-US22886.
XX	
PR	02-OCT-1998; 98US-0102748.
PR	11-MAR-1999; 99US-0123810.
XX	17-JUN-1999; 99US-0139650.
XX	
PA	(KOSA-) KOSAN BIOSCIENCES INC.
XX	
PI	Reeves C, Chu D, Khosla C, Santl D, Wu K;
XX	
DR	WPI: 2000-317716/27.
XX	P-PSDB: NAY84727.
XX	
PT	New isolated polyketide synthase nucleic acid and polyketide compounds,
PT	useful for treating e.g. transplant rejection, uveitis, multiple
PT	sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT	peripheral neuropathy -
XX	
PS	Example 1; Page 82-85; 126pp; English.
XX	
CC	The present sequence represents an NheI-XhoII fragment that encodes
CC	module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
CC	endogenous acyltransferase (AT) domain replaced by the AT domain of
CC	module 12 of the rapamycin PKS (which is specific for malonyl CoA).
CC	FK-506 is a potent immunosuppressant, and acts through initial formation
CC	of an intermediate complex with protein immunophilins known as FK-506
CC	binding proteins. The nucleic acids are used for producing polyketide
CC	compounds. The polyketide compounds can be used as immunosuppressants
CC	to prevent or treat transplant rejection, graft-versus-host disease or
CC	uveitis. They can also be used for treating e.g. alopecia universalis,
CC	autoimmune chronic active hepatitis, inflammatory bowel disease,
CC	multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC	also have neurotrophic activity and can be used to promote neurite
CC	outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC	and in intact animals, they promote regrowth of damaged facial and
CC	sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC	the brain. They can also be used for treating e.g. Parkinson's disease,
CC	Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC	peripheral neuropathies. They can also be used in agricultural and
CC	veterinary applications.
XX	
50	Sequence 4466 BP; 634 A; 1796 C; 1433 G; 603 T; 0 other;

CC	multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC	also have neurotrophic activity and can be used to promote neurite
CC	outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC	and in intact animals, they promote regrowth of damaged facial and
CC	sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC	the brain. They can also be used for treating e.g. Parkinson's disease,
CC	Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC	peripheral neuropathies. They can also be used in agricultural and
CC	veterinary applications.
XX	
XX	Sequence 4478 BP; 632 A; 1757 C; 1465 G; 624 T; 0 other;
XX	
QY	Query Match 31.1%; Score 31.4; DB 21; Length 4478;
Db	Best Local Similarity 67.7%; Pred. No. 7.2;
	Matches 44; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY	37 TCAGCGTGAAGCGCCGCTCACGAGCCTGGGAATGACTCGCTATGGGCTAGAGCTGC 96
Db	85 TCCCCGCGAGCGCGGCGCTTCAAGACCCGCGCATCGACTCGCTACCGCGGTCCAGCTGC 144
QY	97 GCAC 101
Db	145 GCAC 149
XX	
XX	RESULT 15
XX	AAAA14664
XX	AAAA14664 standard; DNA; 4547 BP.
XX	AAAA14664;
XX	08-AUG-2000 (first entry)
DE	Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX	
KW	FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW	immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW	transplant rejection; graft-versus-host disease; alopecia universalis;
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;
KW	multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW	neurite outgrowth; nerve regrowth; Parkinson's disease;
KW	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW	peripheral neuropathy; ss.
XX	
OS	Synthetic.
OS	Streptomyces hygroscopicus.
XX	
FT	key Location/Qualifiers
FT	CDS 9..4535
FT	/*tag= a
FT	/note= "no termination codon given"
XX	
XX	WO200020601-A2.
XX	
XX	13-APR-2000.
PD	
XX	
XX	01-OCT-1999; 99MO-US22886.
XX	
PR	02-OCT-1998; 98US-0102748.
PR	11-MAR-1999; 99US-0123810.
PR	17-JUN-1999; 99US-0139650.
XX	
PA	(KOSA-) KOSAN BIOSCIENCES INC.
XX	
PI	Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX	
XX	WPI; 2000-317716/27.
DR	P-PSDB: AAY84728.
XX	
PT	New isolated polyketide synthase nucleic acid and polyketide compounds,
PT	useful for treating e.g. transplant rejection, uveitis, multiple
PT	sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT	peripheral neuropathy -

XX Example 1: Page 85-88; 126pp; English.

CC The present sequence represents an NheII-XhoII fragment that encodes
 CC module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
 CC endogenous acyltransferase (AT) domain replaced by the AT domain of
 CC module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
 CC FK-506 is a potent immunosuppressant, and acts through initial formation
 CC of an intermediate complex with protein immunophilins known as FK-506
 CC binding proteins. The nucleic acids are used for producing polyketide
 CC compounds. The polyketide compounds can be used as immunosuppressants to
 CC prevent or treat transplant rejection, graft-versus-host disease or
 CC uveitis. They can also be used for treating e.g. alopecia universalis,
 CC autoimmune chronic active hepatitis, inflammatory bowel disease,
 CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
 CC also have neurotrophic activity and can be used to promote neurite
 CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
 CC and in intact animals, they promote regrowth of damaged facial and
 CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
 CC the brain. They can also be used for treating e.g. Parkinson's disease,
 CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
 CC peripheral neuropathies. They can also be used in agricultural and
 CC veterinary applications.

XX Sequence 4547 BP; 658 A; 1851 C; 1437 G; 601 T; 0 other;

Query Match 31.1%; Score 31.4; DB 21; Length 4547;

Best Local Similarity 67.7%; Pred. No. 7.2; Mismatches 0; Gaps 0;

Matches 44; Conservative 0; Indels 21; Indels 0;

QY 37 TCGAGCTGATGCGCGCTCAGAGCCTGGAAATGACATCGCTGATGGGCTAGAGCTGC 96
 Db 85 TCCCGCGAGCGCGGCTTCAAGAGACCTCGCATCGACTCGTACCGCGGCTCAGCTGC 144

QY 97 GCAAC 101
 Db 145 GCAAC 149

Search completed: November 6, 2002, 13:16:46
 Job time : 255 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 : Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_37570_37670

Perfect score: 101

Sequence: 1 aggtctgcagcgtctgcgc.....tgggctcagagctgcgaac 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pat:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	71989	6 AR172664	AR172664 Sorangium
3	99.4	98.4	68750	1 AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6 AR193029	AR193029 Sorangium
5	99.4	98.4	68750	6 AR199551	AR199551 Sorangium
6	99.4	98.4	68750	6 AR199559	AR199559 Sorangium
7	99.4	98.4	68750	6 AR199567	AR199567 Sorangium
8	99.4	98.4	68750	6 AR201097	AR201097 Sorangium
9	99.4	98.4	68750	6 AR208671	AR208671 Sorangium
10	80.2	79.4	16124	1 AX024384	AX024384 Sorangium
11	80.2	79.4	16124	6 AX024277	AX024277 Sorangium
12	49.2	48.7	978	1 AX024375	AX024375 Sorangium
13	49.2	48.7	978	6 AX024268	AX024268 Sorangium
14	45.4	45.0	42603	1 AF188287	AF188287 Stigmatel
15	43.6	43.2	82746	1 AF453501	AF453501 Actinosyn
16	40.4	40.0	66808	1 SAU421825	SAU421825 Stigmatel
17	40.2	39.8	5292	6 AX153793	AX153793 Sequence
18	40.2	39.8	34071	6 AX153790	AX153790 Sequence
19	40.2	39.8	42717	6 AX153789	AX153789 Sequence
20	40.2	39.8	49736	1 AF319998	AF319998 Stigmatel
21	38.6	38.2	30000	6 AX250261	AX250261 Magnaport
22	37.6	37.2	105495	2 AC114475	AC114475 Magnaport
23	35.4	35.0	104326	1 AB070940	AB070940 Streptomy
24	35.2	34.9	33529	6 AR166425	AR166425 Sequence
25	35.2	34.9	105495	2 AC114475	AC114475 Magnaport
26	34.4	34.1	14874	1 AE005927	AE005927 Caulobact
27	34.2	33.9	4615	6 AX153791	AX153791 Sequence
28	33.8	33.5	11890	1 MTCY275	MTCY275
29	33.8	33.5	14980	1 SC4A9	SC4A9
30	33.8	33.5	18614	1 AE007033	AE007033 Mycobacte
31	33.8	33.5	30000	6 AX250262	AX250262 Sequence
32	33.6	33.3	14402	1 AF098795	AF098795 Pseudomon
33	33.3	32.7	1721	8 AF488522	AF488522 Oryza sat
34	33.3	32.7	1784	1 AF411573S3	AF411573 Actinomad
35	33.3	32.7	1791	1 PST249741	AJ249741 Pseudomon
36	33.3	32.7	6564	6 AX069073	AX069073 Sequence
37	33.3	32.7	20679	1 MTCY19H9	MTCY19H9
38	33.3	32.7	32870	1 AE007122	AE007122 Mycobacteri
39	33.3	32.7	107379	1 SHGCP1R	SHGCP1R
40	33.3	32.7	112133	2 AP005254	AP005254 Oryza sat
41	32.8	32.5	43280	1 SFU78289	UF78289 Streptomyce
42	32.6	32.3	20235	1 SPRER3AB	M63677 S. erythraea
43	32.6	32.3	20235	6 AR049368	AR049368 Sequence
44	32.6	32.3	20235	6 AR095529	AR095529 Sequence
45	32.6	32.3	20444	1 SEERYABS	X62569 S. erythraea

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
AF217189	AF217189	Sorangium cellulosum putative transposase gene, partial cds;	AF217189	AF217189	complete cds.	AF217189	GI:7453554	1
		putative transposase gene, complete cds; epothilone biosynthesis						
		gene cluster, complete sequence; putative membrane protein gene,						
		complete cds.						
		Sorangium cellulosum.						
		Polyangium cellulosum						
		Bacteria; Proteobacteria; delta subdivision; Myxobacteria;						
		Myxococcales; Sorangineae; Polyangineae; Polyangium.						
		(bases 1 to 58733)						

Pred. No. is the number of results predicted by chance to have a

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 2015953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1..58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1..992
/note="orfa"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHEAFADFGVPRVLLYDNLKNAVYERHGDAIRFHTLLAT
SADYFEPRPVAVARGNEKGRVERAIRYREGFPEARAVADGLNQTATWYSSAL
DSWVEDRATYRQAQFDDERSVILRHPTPEPDHREVEVEGKTPYARQDLNDYSVP
DTRRTLVYADLSQVRIADGNQIVATHTYRSDMGQIIEOPHIOIRYDDEKRAEHR
GIDRLARARASQATFLRTYAEAGDNVGSILRLQIDLVGAALAEALAEVLENDTI
HGAVRYDIRRSEHRLEPPVSIPTVKEHALVYTPHSLTTIDYLLKQDP"
989..1501
/note="orfb"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTTEKDKRLKSLGGLLACWEOQLADKRWLEVALIERE
RHKRLERLKNKSRVAFAEKPMTPDPSNPKRIDRAVDOLYDSRADLLFEVYTRRD
AOKPLLSLNKAFADMGQVYPAACVYTLVDRLVRAEVIETIASYRLEKAEKELNAT
RTKORRTKKH"
1998..6263
/gene="epoa"
1998..6263
/gene="epoa"
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="Epoa"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRIPEAAEDPIAIVGAGRLPGCVIDISGFWTLLGSRDVT
GOVPAERMDAAWEPDLDAPGKTPYTRASFSDVACEDPASFFGISPRFALLMDPAHR
LILEVCWELNAAIAPSAIVGTETGVLTGIPSPQIGDVAFLATTRSAEHLAAVAT
PSVAGRISYVLTGRCPCVAVDTAVSSLVAVHACOSLRSGECSTALAGGLSLMSP
STLVWLSKTRATLADGRCFAADGAFGRCGCAVYVVKRLSGCARADDRLLAVYRG
SAIHDGASGGLTPNGSSQETVLKRALADAGCAASVGVYFAHGTTLGDPIEIOA
LNAVIGLRDVAITPLIGSKTNLGRHAGSITGLKLVLSLQHQQIFAHHLAAQALN
PRISMGDLRTVTRARTPMDNTPRRAGVSSFGMSGNAHVLEAPATCTPPAPE
RPAELLVLSARTAAALDAHARLDHLETPSQICGDVAFLATTRSAEHLAAVAT
SSEGRALDAAGCOTPGVYRGIAIDSSRGKLAFLTGOGAQTGMRGLYDVNPAF
RAFDICVRLFNQEDRLREYMAAEPASVDALADONTAFOTPALTFPEYALALMRS
WGPEEIVAGHSIGELVACVAGVPSLEDAVFLVARGLMALPAGGMYSTAAPEA
DYAAAVAPRAASVSTAAVNGPQVIYIAGKGPVYHIAAAMAKGARTKALYSHAFHS
PLMAPLELFGVAAVSIVRRPSIVLSMLSGAGTDESSPGYWRVRAAREVVRADG
VYALHAAGLFGFVEGPKSTLLGLVPACLPDARPALLASRAGRDEPVEALGLW

AVGLVSNAGLEPSSGRRVPLPTTQMGBERYWIDTKADDAAGDRARAGADEVEYK
GAVRGDSARLDHPPESSGRREKVEYVAGDAPFLEIDEDEGVLDRLVRYTERSAPG
LGEVEIADVAAGLSEFNQOLAGVAVPDLPKGNPPLLEIGGCACATYVAGVGGVAV
VGOPIYALISAGAFATHYVTSALVLPKQALSATPAAMPAYVLTIAVADLGIAPLOP
GERVLIHATGCGVGLAAVQMAOHVCAEYHATAGTEKARAYLESGLVYVRSRGRFV
ADVRAATGEGEVVNLNSGLDLKSNLSEHREVELEKRDYCANQOLGRFELR
NLSEFLVDLRGMLEPARVRALEFELLGLIAAGFPPTPIATLPIAVVAARMAQ
AQHGLKLVLTLDGPEVOAIRIPTHAGAGSTGDRDLDRLSAAPARAALAEALFRTQ
VSOVLRTEIKVGAALFTRIGMDSLMVELRNITIASIKLSITTFISTSPNIALT
ONLDAIATALSLEVAENLRAGVQDSFVSSGADQDMETIAL"
6260..10492
/gene="epob"
6260..10492
/gene="epob"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQILNLEFHOGVTKLAADGERLQIAPKNAIPLNLTARISEH
KSTITITMLRQRLPASTYVPAPEKRVPPPLDIDQSTYGLGRGATTVSGCHAREYD
CTDLVVARLSRAFKRVVAHMDLRAHTLPDMQVLEPVVDADIEITLDLRGDRSTREA
RLVSLRDAMSHRIYDTERPPYVAVRLDEQTRLVSLIDLINVDLGLSITIFEDWL
SFYEDPETSLPYLESLRDYVLALESRKSEHQRSMQVWRKRVALEPPLPKAD
PSTLREIRERHTEOMLPDSMSRLKORVGERGLPTPTGILAESEVIGRMSASPETL
NITLFNRLPVHPRVADITGDFTSWVLLIDITRDKSFEPRAKRIOEOLWEMADVDV
GIEVOREARVIGIORGALFPVYVLSALNQOVYTSIQRLGTPYVTSQTPOLLDR
QLYERHGDVYLAWDIVDVFPPDLDDMLEAVAVLRRLTEPSEQRCSIPPAOLE
AASANENLSLSEHTLGLFAARVEQLPMQILVARSAKTLYLELSRRSRRLGARLR
EGGARPNLVAAYMERKGEQVAVYLAIVESGAAYPIDADPAEIRIHLVLDHGEVLY
LTQPLDGLSWPGLPIRLVSDADEVGDGLDPMPIQOTSPDLVAVYVYTGSGSTLPK
GVMIIDHGVNVTIIDINREIFLEGDVRVLAISLSLEISYDVDEGIIAAGGTIVPDA
SKLRDPAHMAALIEREKYTVMSYALRMLVHSESGPDSIASRLSLSGCMIPY
GLPGLQAIRGCVYISIGGATASITWIGIPKRVNDVSMASITPGRLDKQTRVLDP
EALERPVAWVPGQITIGVGLALGRWDEEKRNLVHPETGERLYTGDGRVLP
GNIEFMRGREDNQIKLGRYVELGEIETEELKSPNVNDAVIVPVDGANKLLAVYVP
EGTRRRAEQDASLKTERVDARAHAAKADGLSDGRVOFKLARHGRDLQKPYVDL
TGLVREAGLDVYARRRSYRTELEAPIFEVEGRFLSGLSVEPQALPFRPVSAG
SYPVQTYAAYAKSGRIEGVDEGFYVYHPEHRLKVSQHGIERGAHVONPDVDEA
PELLVGRIDATIESLYGSLSRFCLLEKGYAQLMEQAPSCNITCVCGQPDVEQYR
PVLDIRHSDVYVHGMVGVDPRQVCTTLDGSSPRATRGAPRGKHQHPADILRD
FLTRKLPEYMPVEVEDALPTNSGVYRKMALRERDTSPRSHSGTAPRADLEEI
LVAAVREVYGLTEVGLQOSFDVLAITSIHVMSRLQKLRDEIATITELQYPLGS
IASGLRKSXKDQERPNQDREYARKGRRRS"
10639..16137
/gene="epoc"
10639..16137
/gene="epoc"
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEBOESSAIAVIGSGRPPGARLDDEFMRNLROGTEAVORSEQ
ELASAGPAAVLVDPNVYRAGSVLEDPDPAAPGCTSPRAEALMDPHRLEMECAWE
ALENGYDPTAEGSITGYVAGANMSSTYLSNLHEHPARMRPGQTLIGNDKQILATY
HVSYRLNLRGPSISVQTAGCSTLVAVHACNLSLHEDREDMALAGITVIRIPRAGYV
AEGGTFSPBGRAPDANANGTIMNGGVVLLKRLDLALSGDPRVAVIIGSATND
GARKIGTFAVSEVGOAQMELANGIAYEARIOYIEPHGTTLLGDALIEETALAEV
DBDASTRSCASIVKTIIGHLESAGIAGLIKTYVALLEHQLDPLSPLESPNSIDF
AASFPYVTSIKDNWGTSTPRRAGVSGSIGGTANVYLEERPAKLPAAAPARAEAL
FVYSKSAALDAARLRDLQAHQGLSGLGVAFLATTRSPMHRILAMAPSAEAL
REGDLAARGQTPPAVGRKCSPGNVVYVFFPGQGWAVGMMGLALMRGVPFIAL
SACDPAIIEAGMGVSLAEADANGSQTLEIDVQVYVLAQVAGMGALAEAGVGPV
VIGHSGVEVAANHVAGALSLEPAVALICRSRLSLGCGCMATVETLSLEBAALR
GYEDSVVAASVNSPSTYLSGPRVALITGSLSNAKKYCGGVYKDVASHPDQPLR
EDLLAALGLRPAARAAVPMRSTVYTGAMYAGPBLGANTYMNRLRQVRYRABEVYQALOG
GHGLFEVSPHPILTTSVEEMKRAAORGAAVGSIRQODEPAPALIEGLALMAGYIP
VPMGRIFPAGGRVPLPTYPMQREERYTEAPAKSAAGDRGVRAGGHPLLBMOULTST

QSTRLEWETTLIDLRLPMLGDHROGAVNFEPAALYLEMAISSGAELDGPLOITDVA
LAELAFAAGDAVAVOYVTTTEOPSGRLOFOJASRPAAGHAFSPVHARGALLRVERTE
VPAGLITLSAVARALQASMPAAATTAETEMKLOQGPAPOGIAELMRGALLGVRLE
DAGSAEIRLRLPDLNCFQVGSIFRAGGAEATPWVEGSLRLQRPGLRCH
RVNNGHQTDPDROGADFWYDSSGVAVAEVSIAVORLPJGVRRREEDNPLLEWEP
AAVGTAVNNGRMILLGGGGIGALRMSLEAGHAAVHAASHTSMAAGVALLAKAF
DGOPTAVNHLGSLDGGELDPGAGALDAPARSADVPALDPALVARGDSDVLMVY
OALAGMGEFRAPRLMLTRGQAVAGADVSTQAPLIGIGRIVTMEHRAFCARVDL
PTRPDGLGALIAELADDAEAYALNGERCVARITYRROPTETPRRISCYPTDVT
IRADSTLVTVGGJLGLSVAGMLAENGAGHLVYVSGSAVSQRAVAALIEARGAR
VTYAKADVADRAQLERILREVTTISGMPLRGVHAAGLIDDLMOQPPARPKYMAP
VOGALHHLHALTREALSEFVLYASVGLISPGGAGNYAAATFEIDALAHHRRAKAF
LSYDMGJFAEYVMAAOEDRGARLVSRGMSRLTPEDEGISAATRLLESRGVQWGPVN
PLWVELYPAASSRMLSLRYTARASAGGAPGDGLLRLRAAEPAARSGLLEPILR
AOISQVRLBPGRKEVDAPLITSLGMSLMELENNRLEAMGITTVPATYLTMTYTVAA
LSGHLAREACEAAEVESPHTTADSAVEIEEMSDTLTQLIAFKKALT"

gene
16134 .37907
/gene="epod"
16134 .37907
/gene="epod"
/note="type 1 polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=1
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQQNPILKQAIIITQRIEERLAGIAQAELETERPIATV
GIGCFRPGADAPAEFWELLDAERDAVOPDLRRMALVGAVEAVPMWAGLTPIDC
FDAFFGISPREASLDPOHRLLEWMEGLDEDIPRSIDGSRTEVFVGAFLADYA
RYVARLPREERDAYSATGNMLSIAGSLSTLIGLOCLVYVTCSSSLVAILHAGS
LRAGESDLALAGVSTLSPDMMEAAARTQALSPDRCGRPDASANFMEGREGGLVY
LKRLSDAQRGDRIMALIRGSATIHNDRSRGTLPRTVLAQETVYREALRSNHYAGAY
DYVTEHGTGSLDPIEVEALRATVGPARSQGTCEVAGVNTNIGHLEAAGVGLIK
AALSTHERGIPRLNFTLNPRLRLREGSALALATEPVWPTDRPRAGVSGFMSGT
NAHVLEEAPEAVELMPAPERSABELVLSGSEGALDQAARLEHDMHELGILGY
AFSLATRSAMTHRLAAVATSRREGILAAVLSAGOGPAGARCIASSSRKILLFT
GOGAOTPGMGRLCAAMPAPREARDRCVTLFDRDLDPRLPVRMAEFGASLILDDT
AFQORLFAVEYALTAAMRSNGVPELLVGHISIGELYACVAFSLDEGRVLYAAG
RLMOGLSAGAMVSLGAPEVAEVAAPVPAHAWSTAIVNGEVOYVAGVEDAVAIATA
GFARGVTRTLVSHAFSHPLMEPLMEFGRAVAASYTRRPSVSLVSNLSGKVTDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 5,5e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGTCGCGAGGTGCTGCCCTCCCGAAGCAAGCTGACGTGATGCGCGCTCAGCA 60
|||||
Db 37570 AGGTCGCGAGGTGCTGCCCTCCCGAAGCAAGCTGACGTGATGCGCGCTCAGCA 37629
|||||
QY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGGCGCAAC 101
|||||
Db 37630 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGGCGCAAC 37670
|||||

RESULT 2
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155.
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Reconfirm methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
Location/Qualifiers
1..71989
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5,4e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCTTCGAGGTGCTGGCGCTCCCGAAGCAAGCACTGACGTGATGCGCGCTCAGCA 60
|||||
Db 37570 AGCTTCGAGGTGCTGGCGCTCCCGAAGCAAGCACTGACGTGATGCGCGCTCAGCA 37629
|||||

QY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGGCGCAAC 101
|||||
Db 37630 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGGCGCAAC 37670
|||||

RESULT 3
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain so ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE .
ORGANISM Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Mycococcales; Sorangiineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Stratanow,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
agents. Biol. 7 (2), 97-109 (2000)

JOURNAL
MEDLINE 10662695
PUBMED 20130945
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA

FEATURES
source
1..68750
Location/Qualifiers
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=1
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSARVLDDGGRATERHVLAENGIEDLRALREH
LRIOGESPFRCHKGLDITVLELHAHDPLASISPHHARSRLHPWPTEDAMLVGPRAY
RWLAKAPGCPRLREYERERKARTAQERARLWLAAPCPAPDLPREDNANGLPPLP
MSPEVAERLRALRSYATPELACALALAWLGTGPGWGPAYEMLEPNLGLGLPT
AIAAASAPGTSEALRGAARLFAEMEVVSSKSGINIPALMERLRTIYRAMGNADN
LSRPERAIAAEVRLRAOPAPPAAGAGLAVAGSSGRLGLVTGDALYSGDGD
IVMFOGRLISVVVLACNDPEPELAPLSQWLFAVHANNAGTISKVLTEGSPLIYMAR
QARMSLYHARGFAMVNOAMVPPREGAPVYVORSTIMEHEHTPPCLHEPAGSAS
LACDEHLVWCELSAGRLERLRRHRRHRRGASRAYLGEHIAATWTPLFLNATHVL
WADPDRAALIGVDKRTVEPIVLAETHRPAHVASEDDRDIFALIGDPSRDWHEHRL
SGASTVADYORQLMDRDPDWVLRGLFTTNDRIILTIARS"

CDS
1..68750
Location/Qualifiers
1..68750
/organism="unknown"
/transl_table=1
/product="unknown"
/protein_id="AAF26905.1"

/db_xref="GI:6724239"
/translation="MGALISVAPAGCALGAEERGPGODAGAGALAPAREVMAEVA
ACOMPAAVLVARGDVDAVGTALGSAAPMRDITIPRIASMTKAVATAMVLE
EGKLDDSPVDRMLPELAPKRVLARIDGPIDEIVAPERPITVADLMTPTMGEGISDA
SSPIORADELGLVNAQVPKMTPHGDEWIRLIGCTPLMHQBAQMMYNGSIOCVL
VGRADQGFDAVYERERILAPLGRDITDIFVPADKARFAGCGFTIDEQTEKTRMD
GAEASVAPSPAPSGAAGLVSTVDYLLPRLMNGVHEGRLLSAGSREMTADHL
TPAOKAASSFPFGFETTHQMGVMAVTAADVAVSEVGRYGMGGGFTSWINDPGREL
IGIVMTOSAGFLFSGALERFRMSVYATESA"
3415. .5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGLTREVLIISVLTALITVYARASGELARLROPEVIGELFG
GVYLGPSVYGALAPGFRHALFDEPAVGVLSTIGTIGALLLIMAGIEVDVGLRKEA
RPGALSALGAIAPPLAAGAAFSALVLDRLPSGLGIVLSTVAVSVIAKVLIERESM
RSSYQVTLAAGVSEVAAWLVAMTSSSGASPALAVARSLASGLFLFWLVGR
LTHLARVYADATRVSGQVSLVLTFLAALTORLGLHPLGAFALGVLNAPRT
NRLDGVOTLVAGLFAVPFVYLAGRVVDSQLRPFAANGVYLLALTTPDAKKVAVAA
LGARLGLRSEPAALVAGLNMGCTDLIYAYVGLGILSNAYINMYAVVLYTTA
SPALLIMLEKRAPTOESARLEEREPARARATIPGVERILVPIVAHALGEFTDIES
IVASKRIGETVDITELSVQQAPGPSRAGEASRLARLGRVIGIRORRELGS
IQAILRASRDHLVIGARSPARAGMSFGLQDAIVQRAESNVLVGDPPAAERAS
ARRILVPIIGLEYSFAADLAHVALAMDALVILSSAQTDPCAAYVWRDRESRVAV
ARSVDVEAVRGRGLGVRSRVYHVGAFSDETRELRANAPVDLVLGCTYDHGPIGLRL
YLSYVESVYVRSRVPVALLVHAGCTREQVR"
complement(5612. .5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGRTRCAIAECYIPGSGNGPEPQMTSHEACULNASDRDA
QVAITYVEFDRDPAGCPYRVTVPARRTRHVRFNLDTEPEPIPRDTYASVIESDAPIV
QVRLDSRQAEMLSTIAYTDRE"
6226. .6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVDYGRFRGRIRIGHTIGLASMALAGCGPSEKTYOCTRIAP
GADARTADVDDPAATTRLAVDVYHLSPPERLEAGSERVYVWRPSPBWRVAVLD
YNADSRGKLAETVYANFELITAEKQSSPQSPSSAAVIGTFSVG"
7610. .11875
/gene="epoa"
7610. .11875
/gene="epoa"
/note="EPoS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRIERAEDPIAIVGASCRLPBGVIDLSGFWLLBESSRPTV
GVPAERMDAAAFDDPDAPCKTPTVTRASFSDVACFDASFFGISPRRALRMDPAPHR
LLEVCWMALENNAIIPSAIVGTEGVFIIGISEVEALPQNTASAEIDAHGICLTMT
PSVAGRI SYALGRGCVAVADTAVSSSLVAILACOSLRSGCSFALGAGVSLIMP
STLWMLSKTRLARDCRKAFASEADGPGRGECACAVYVLRKLSGARADDRILTAIVGR
SAIINHGAASSGLTVPRNGSQEYVLRKALADAGCAASVGYVEAHGCTTGLQPIETQA
LNAVYGLGRDVAIPLIGSVKTNLGHPEYASGTTGLKVVLSLQHQTPAHLHLQAPLN
PRIWGLDRLTYTRARTPMDNTPRRAGVSSGMSGCTNAHVULEAPATCTPPAPE
RPAELLVLSARTASALDAQARLRDLETYPSCQLDVAFSLATTRASMEHRLAVAT
SRGELAAALDAAAGOTSPGAVSIADSSRGKLAFLFTGGAOTTLGMGQVLDVMAF
RFAFDICVRLFNQEDRLPLEVYMAEPAVDADLDQTAFTQPALTFPRYALAAIARS
WGEPELVYGHSTIGELVAAQVAVSLEDAVTLVARGRLKQALPRPGAMVSLTEAEB
DVAAAAPHAASVINAANVAPQVVITAGAGOPVHAIAAAMAARGARTKLAHSHASHS
PLMAAPLBAFGRVAVESVYRPSIVLVSNLISGACDDEVSSPGYVWRHARREVVRFADG

VKALHAAGAGTFEVENGPKSTLLGLVPCAMPDAPRALNASSRAGDEPATVLEAGLW
AVGGLSNAGLPPSGGRVRLDPTYPWQRERRVYIDTKADDAAGDRRAPAGDEVEEG
AVGAGDGRSARLDPPESGREKVEADPEFRLEIDEPVLEVLVRYLVERAPAG
LGEVEIAYVAGSEFNDVOLQAMPDDLPKGNPPLILGGSCARIVAGVAGVGLV
VGOVPVLAAGAPATHVTSAALVLPAPRALSAIRAAAMPVAYITLAWADLBIATLOP
GERVLIHAATGGVGLAANVQAOHVAGNEVHATIGTEPKRAYLSLGVRYVDSRSRBFV
ADVRAWTGGEEVDVNLNLSLGLIKSEFNLNLSHCRFELGRDGYADVQDLRLPFLR
NLFSFLVDLRGMKLEPARVARALBELLEGLLAAGYTFPPPIATLP IARYADAFRMAQ
AOHGLKVLITLGDPEVQIRITPHAGAGPSTGDRDLDRASAAPARAALAEFLRTO
VSQVLRPEITKGAEALFRLGMDSLMAVELNRRIEASIKLSTTFPLSTSPNALLA
ONLIDALATALSLERYAENLDAQVQNDPVSSGADQDWEITNL"
7640. .8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236. .10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529. .11428
/gene="epoa"
/note="Region: enoyl reductase"
11549. .11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872. .16104
/gene="epop"
11872. .16104
/gene="epop"
/note="EPoS P: contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQLELEHOGIKILADGERIQIQAPKNALNPLIARISEH
KSTILMLRQRLPAESIVAPAEERHAPFLTDIOESYWGARGATVPSGIAVREYD
CNDIDVPRLSRAFRKVVARHMDLRAHTLPDDMQVLEPKVADAEIIDLGRDLSREA
RLVSLDAMSHRYDTERPPLVHVAVRDEKOTRLVSLIDLINDIGSLITIFKDWL
SFYDEPETSPLVLELSYRDVYALTESRKSSEAHOSMDQWKRIIEPPLPKAD
PSTIKETPRRHTBOMLPSDSGRLKRVBERGCTIPPGYILAEFSEYIGRMSAPRTL
NITLENRLPHPRVNDITGDFTSWLDLIDTRKDSFEQARARIDQMLEANDHCDVS
GIEVREARVVLQIORGALPEVYLSALNOQVGYSTLQRLGTPVYTSQTQLDLH
QVYEHGDDVLVAMDYDGVPEPDLDDMLAEVAFVLRITEEPGEQVSCSPAPQLE
ARASANATNALISEHTLHGIFPAFARYOULMOEAVASRATLYEELSESRRLGARLR
EGGARNTLIVAVMEKQMEQVAVIAYLESAGAYVPIDMDLPAERHYLHDGEVKLV
LTOPILDKLSNPPGIGIQLVSEAGYEGCGDDPPMPPIOTPESDLAYIYTSSTGCLPK
GVNIDIRGAVNITLIDINERFEIGPDQVYALSLSDLSVYDFGLAAGTIVPDA
SKLRDPAHAEELIEREKVTVMNSVPALMMLVHEHGRDLSLRSLRSLSGMDIPV
GLPGELOAIRPQSVISLGGATASIGSYGVVAVNDLSMASIPYGRPLRNOTFVLD
EALERPVAWPGOLVIGVGLAGYWRDEKRRKSLVHPETGERNLKTKGDLGRYLPD
GNTEPMGRDNQIKLRGYRVELGETEETLSKPNVNDAYIVPQDANKLLIAYVP
EGTRRAAODASLKTERIDARAHAAEADGLSDGERVQKRLARHGLRDLCKPYVDL
TGQDPRAGLDVYARRRSVTRFLEAVIPVEHGRFLTSCISVPEGATLPKRYPSAG
STYPQVTVAVYKSGRLEGVDEGFYVYHRLKLSDIGERGAHVARNPFDVDEAA
PVLDFYGRIDAIESLYGSSRECLLEAQMLMEQOCSNICVCGPVGQNFQVR
FYLDLRHSDVYVYGMGLGVRDPROVQCYLGDSSSRRAKTPGAPGRGROMLDRD
FLRTKLPYMPVYVEYLDALPLTSSKGVKRAKLRKRTSSPRHSGHTAPDALDEI
LVAAYVEVIGLEVYGLQDSFVDLGATSHIYMKRSLQRLDRREIATITELFPNLGS
LASGLRDSRDDLDQRPNMQDRVEVRKGRRRS"
16251. .21749
/gene="epob"
16251. .21749
/gene="epob"
/note="EPoS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match

98.4%; Score 99.4; DB 1; Length 68750;

Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 4
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 5
LOCUS AR199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 6
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 7
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothiliones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 60
|||||
Db 43182 AGGTCGACAGTGTGGGCTCCCGAAGGCAAGCTGAGTGGCGGCTCAGCA 43241

Qy 61 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 101
|||||
Db 43242 GCCTGGGAATGAGCTCGCTGATGGGGCTAGAGCTCGGCAAC 43282

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCGCTCCCGAAGCGCAAGCTCGACGTGATGCGCGCTCAGCA 60
|||||
DB 43182 AGGTCGCGAGGTGTCGCGCTCCCGAAGCGCAAGCTCGACGTGATGCGCGCTCAGCA 43241
|||||
OY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
DB 43242 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCGCAAC 43282
|||||
RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.3e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCGCTCCCGAAGCGCAAGCTCGACGTGATGCGCGCTCAGCA 60
|||||
DB 43182 AGGTCGCGAGGTGTCGCGCTCCCGAAGCGCAAGCTCGACGTGATGCGCGCTCAGCA 43241
|||||
OY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
DB 43242 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCGCAAC 43282
|||||
RESULT 10
LOCUS AX024384 16124 bp DNA linear BCT 15-SEP-2000
AX024384 16124 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
Location/Qualifiers
source 1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 79.4%; Score 80.2; DB 1; Length 16124;
Best Local Similarity 87.1%; Pred. No. 6.9e-10;
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 1 AGGTCGCGAGGTGTCGCGCTCCCGAAGCGCAAGCTCGACGTGATGCGCGCTCAGCA 60
|||||
DB 5947 AGGTCGCGAGGTGTCGCGCTCCCGAAGCGCAAGCTCGACGTGATGCGCGCTCAGCA 6006
|||||
OY 61 GCCTGGGAATGACCTGCTGATGGGGCTAGAGCTGCGCAAC 101
|||||
DB 6007 GCATGGCATGACCTGCTGATGAGCTGAGAGCTGCGCAAC 6047
|||||
RESULT 12
LOCUS AX024375 978 bp DNA linear BCT 15-SEP-2000
DEFINITION Sequence 73 from Patent DE19846493.
ACCESSION AX024375
VERSION AX024375.1 GI:10184579

REFERENCE	1	Myxococcales; Cystobacteriineae; Cystobacteraceae; Stigmatella.
AUTHORS	1 (bases 1 to 42603)	Slitakowski, B., Schaiter, H.U., Ehret, H., Kunze, B., Weinig, S., Nordtek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S. and Muller, R.
TITLE		New lessons for combinatorial biosynthesis from myxobacteria. The myxochinaiol biosynthetic gene cluster of Stigmatella aurantiaca DM4/3-1
JOURNAL	J. Biol. Chem.	274 (52), 37391-37399 (1999)
MEDLINE	20069734	
PUBMED	10601310	
REFERENCE	2 (bases 1 to 42603)	Slitakowski, B., Schaiter, H.U., Ehret, H., Kunze, B., Weinig, S., Nordtek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S. and Muller, R.
AUTHORS		
TITLE		Direct Submission
JOURNAL	Submitted (20-SEP-1999)	NCBI/NC, GBF, Mascheroder Weg 1, Braunschweig 38124, Germany
FEATURES		Location/Qualifiers
source	1. 42603	/organism="Stigmatella aurantiaca"
		/strain="DM4/3-1"
		/db_xref="taxon:41"
gene	1. 834	
		/gene="mtaA"
CDS	1. 834	

gene

CDS

```
/codon_start=1
/translation_table=11
/product="MtaA"
/protein_id="AA119809.1"
/db_xref="GI:6635394"
translation="MPTSSPALPLLKLPPDEVHWIWEPERITEPGILLESYRALDPG
ERKQKQVEFERHRLQIYLVSHALVRLTISYAVAPAPAFSANOQRPRIREEKMG
LRNLSTPDDMCALCAVARQVDGADVEDDEREPEYIADSPAPAVASRALPYSG
QRRFPDYMLTKKAYIKARQMGISLIDPAPFYSOGLRISRPDLNDEPOMQDV
RFRPSQRHAALAVRRSEAPLTVRQRTYPLDDDAFAELISBERIOLPLRRMPQVG
G"
1388. .13399
/gene="mtab"
/note="PKS"
1388. .13399
/gene="mtab"
/note="PKS: Mtab loading domain and module 1: polyketide
synthase domains: ACP, KS, AT, Ar, Dh, ER, KR, ACP; Mtab
module 2: polyketide synthase domains: KS, AT, Dh, KR,
ACP"
/codon_start=1
/translation_table=11
/product="Mtab"
/protein_id="AA19810.1"
/db_xref="GI:6635395"
translation="MSTEHGVGVGNSTGOLRDLWVOOVASQVLEAHVITPEPFS
LGGVSFATRLIAELARWLMEPLTLVWHNPIDIFARLSCAEPRAAMHAGPPA
ANDGAPRIAVLGACRFKPKAPDAKAFRWLOQERDCVTEVPSNRMDASITDPAQTP
GKIVTRGAFILDSVDREDPVYFGISPREAGQLDPQORIMLELSEWLEDGIAPVLR
GKSTGVETGMVWRDYADLOITINHAAPLHTCTGAGIITANRISYALGEGPLAIDT
ACGSSVVAIHEACRSRLRSGESSLALDAGVQNLNTPREPMVASGALAPDGRCTEDA
RACGVRGVEHGGVLYIKPLSRALADDAATCYLRGSVNNNDGASNLTAPNACROBL
LRDAYVAGMDPAQVHYVELHGTGTPGLDITEGALGAVIGARREARPLHVSQVTN
IGLEHAAGIASYIKVAALSTRHOQLPASLHEDEENPEITPDDILLRQTOIAPWAPK
DPAVAGISAFEGFGTNCVHVEEHRGSRGSRAPAAVSPHAGSRHRTALFEGASA
WMMGMRGLGAEPVRFARREECDAARFPGGLMSLEVALLRVRSREDEGRQPLVL
AFQISIALLMRSWGIPEPSATVGVSEBMAAOVAGILITIEVALVHYIQOVGAGV
GVYMAVAVLGSGAEVARIAPYGGRIIASENGSRSTGLGEAALADOFETASMTAGIF
ARVLEIDFYAHIPQVBAIKRIFLOSCPVLRSPGRVPMVTSVAGFVDGDTGAEYFA
SNRREVPRTAMRITLLEGGYVDLVNPNPITHEHYLENVKGTASAEPTIASMRGE
PSYEGIHRALITQSQAGCPYRMERYVPGTAGSSAPYRTFLPAAKARTBALRQADR
IAOULRRDAASSLTIDLCYTNGARKSHHEHRLVAETRALLEGALAFAGELRPNS
QNGWRGEAEPLGFEVSGGSPFGSSVOLFEESPESKAFKESERIFARQATASILE
ALHAEESQHPRTETLEOPLLEFQVALALWESWGRIPDAVYHSGVLEISAAHVAGV
```


REFERENCE 1 Actinosynnema.
AUTHORS Yu,T.W., Bai,L., Clade,D., Hoffmann,D., Toelzer,S., Trinh,K.Q.,
TITLE Xu,J., Moss,S.J., Leistner,E. and Floss,H.G.
JOURNALS The biosynthetic gene cluster of the maytansinoid antitumor agent
MEDLINE ansamitocin from Actinosynnemapretiosum
22056096 Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7968-7973 (2002)
PUBMED 12060743
REFERENCE 2 (bases 1 to 82746)
AUTHORS Yu,T.W., Bai,L., Clade,D., Hoffmann,D., Toelzer,S., Trinh,K.Q.,
TITLE Xu,J., Moss,S.J., Leistner,E. and Floss,H.G.
JOURNALS Direct Submission
Submitted (28-NOV-2001) Department of Chemistry, University of
Washington, Box 351700, Seattle, WA 98195-1700, USA
FEATURES
Source Location/Organism
1. .82746 /organism="Actinosynnema pretiosum subsp. aurantlicum"
/strain="ATCC 31565"
/sub_species="aurantlicum"
/db_xref="ATCC:31565"
/db_xref="taxon:42198"
misc_feature <1. .82653
/note="cluster I of the maytansinoid antitumor agent
ansamitocin biosynthetic gene cluster"
gene <1. .1678
/gene="asm37"
CDS <1. .1678
/gene="asm37"
/codon_start=2
/transl_table=1
/product="unknown"
/protein_id="AAM54115.1"
/db_xref="GI:21449383"
/translation="TVPLVVALATGVLAPASAAITSPVTSTANTVTEVSGT
PASFVRVVDVDRSAATGYAOGIGADHLENGSLYRHSGGAMTQVATVPTASGV
ARMTVERADLGEOSAPNDVYFOVEAPVETSAVTHRTYTAGGSVAYOSSANVP
PERGLYHHTGDCADLSADLRRTGEGITLAMCYFYLAEHRSSTPAOTLDRLO
QFDAARYAGCKLVIRFAYTASESESDAPKREYLAHIIQALPLANADYVAVAGPY
GAWGEMTYTKFNGAGVYASATDQAKRAVNDKLLTALPADRAVOLTFPKERALYGGS
STRIGHNDFCLNDYDPTGYEDPATEPYLEATRTLPMGETCAVNPPTTDSATLP
AELSRFMHTYLNIDYQODVLRAMSAOGLAEVATKLGRTLRDSVFPPTTDSATLP
VRVNLNDGYAAAVNORTVNLRTNTTGALTRLPLTADPTMPLPGARALQDAQIP
SLATGOYALLLELADPALPGRPEXSIOGTANGLDARTGLNDKQITTVS"
1825. .2142
gene /gene="asm36"
CDS 1825. .2142
/gene="asm36"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAM54114.1"
/db_xref="GI:21449382"
/translation="MVTASAVLCARACGAPRPAGATARMFCSYACKQAAYRLRRA
GSASHEHEDEQLREALLALVARARTTEAPRPMSAQATALVGLARVVELAAADAGV
SGC"
gene complement(2129. .3307)
CDS /gene="asm35"
complement(2129. .3307)
/gene="asm35"
/codon_start=1
/transl_table=1
/product="bicyclomycin resistant protein"
/protein_id="AAM54113.1"
/db_xref="GI:21449381"
/translation="MAVGKADSAVGMPLALLIYALTGTTIIMSVSPFDIARELG
VGCPAYOLVTNLNFGCLGALCHGPLSDRGRRPVLGSAVHLVGLGALAPTTG
LLTARFVVGIGASAAVIVFVIIIGDRYGRPALRMYGIVAAASALMTCQALIGAT
NRVAGRGNTLANCATLALSILVFLPFRTRARATITARSYLADYRRLGSRATL
AASLTFPSLFRATLVFTAASPFYLTAAFGVGTGALHLLVVAASFVPSFAFGIIP
LLGGPAGAVRWGLGAAVVGVLGDLIDGPAVAGVALLFCVGFVAVCPVVFEDSMY
FPFLAGAASSATMGLRALLVLTLLTALAGALSTGPGFAPAAVMAAGAVAVVLSLSPR
"

gene complement(3450. .4064)
CDS /gene="asm34"
complement(3450. .4064)
/gene="asm34"
/codon_start=1
/transl_table=1
/product="transcriptional regulator"
/protein_id="AAM54112.1"
/db_xref="GI:21449380"
/translation="MARNPERRAALDAAYEVLAGECARGLTRANDQRCYAGRA
SNYFRAGREILTGAGERYQRLDQAPLGETPDGRDRSYTELHAHADVSAFPG
FIALLELEAARNPELEKRELVLARVADLRNRAHNAASGLPGGTYVALLMLNL
ILERTLPDLTLDHQRDYALVDRLIACQPEBAGSPFDESGR"
4151. .4723
gene /gene="asm33"
CDS 4151. .4723
/gene="asm33"
/codon_start=1
/transl_table=1
/product="dihydrofolate reductase"
/protein_id="AAM54111.1"
/db_xref="GI:21449379"
/translation="MRKLVPFAASIDGVIAPEGEHDFYTAEPMEHLRAEYPTT
PTHYRPLVGMPLDTPNRRYDVLWGRGSYQPALDAGVSSPAHLROYVSRSLDADP
AVALVRADPVLVLELREKREDLDVWLCGADLAGQLPEIDELVTKTYVVAAGGTPL
FRAGFAPRAPRPVYRAFDHGGVYTTARA"
complement(4854. .5057)
gene /gene="asm32"
CDS complement(4854. .5057)
/gene="asm32"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAM54110.1"
/db_xref="GI:21449378"
/translation="MWITEQARIPARRSSGRSNARCAVWMPBPQPSATRADGPS
SANDRSRSGRDPRTSLNSA"
5415. .6461
gene /gene="asm31"
CDS 5415. .6461
/gene="asm31"
/codon_start=1
/transl_table=1
/product="RNA polymerase sigma factor"
/protein_id="AAM54109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRRPVMGVTEAFETWETDEDIKLAATLRDAEA
ANADSTRAYLKAIGRVRLTADQEVLELARVEGLYAEHLDELPPAARGRRPEYLR
RVVRDGORAKDHLLEANRLVYSIAKRYTGMAILLVIOGNGLVYAAVEKFPDYAKG
FKPSTVATMWITROATRLADOSRTIRLPVMAVYVVKIAREIDLHRRNREATAEE
VAVELDYTPPEKVELROHARTPVSLDGTIGAEATSYGDLTADENAYALPENGSAAL
SADLRVAVLETLTPREAGVYRMRYGPPGVLHTLDEGVYGVTKERIKOLESRMAKL
QHPARSRGLEGYLD"
complement(6501. .9518)
gene /gene="asm30"
CDS complement(6501. .9518)
/gene="asm30"
/codon_start=1
/transl_table=1
/product="cytochrome P450"
/protein_id="AAM54108.1"
/db_xref="GI:21449376"
/translation="MVAITGTRIPGKPLPYGNLLDVLTSLDLPDVDFLDRCHREHG
ITVLTFAGQROVFASSHELVARKCSDSWGAIVPALAEQVDFAGDGLFARGGEPM
GKAHRLMPAGPTAMRDRHPPEAMLDIEQMLVRRMRGPPDRIDVADMDRLTDLTA
LCAFGARFNSFYRDRAPHFVDAMRVSLVEAGERLRLPGVOPFLVGNQRRDIAIM
NRIDGTYAAARAALPAGRPDLDLEMLTCDPVTGELSARNTYOLAFPLIGHET
TSGILSPAVHRLIAPHRYLAKRAVDGVCVGRAPAREDLARLDYLGQVARETLRLP
TAPFALAPDEPRLAGHATGAGEPVLYMLPTLHRDRAVRMDPVFPDERAPRMBE
IPACAMPEFGGARACIGRPALOEATLVIALVLQRFDLADLADPDHRLTIKQTLILP
DSLIVRARPPRADRGATAYTVPVHOVPATHRGTPLHYVYGSNGSGEGLARTIG
DGAARGMATSVAPLDDAVRALPAGSPVIVSSSYNGAAPDNAHFVWMLTQDGPDLG
VDVILVLCGNDWMSATYGVVPTLIDEMAAGAARLRLEBRGATDARADFFGDMERYEP
"

LMP1LAEQGVGEVGEIPRVEVSDAAAGLDGLASAVVLENEELVGRPGAGSKRHL
 LR1PGTSIRGDDYLSYLPQNHPIQDILVRAVVRIGTAEVRYVSSAPGILVYGRAL
 RVDELTLTQVDSLAPAGASVVAFLARCCSPREARLATAATGATLLELLEPFGCAND
 LALALELPAPTRLYLSISASAEEDCAEALTYVTVTSGITLSRVRGDRAVGLIS
 PPSRPRPPDNTVPVYLIAGGILAFKFLRARAALDEGPGALLILFGCRGPELID
 YAEERAAALGMLVEVRAVSRHDDGEVRHVOHRLMORRDRVRELVDAGARVYLDCGATIR
 VGPAAVEVLGRIGRPGAGWDLARAGGRVATDVF" 9664 . 10245
 /gene="asm29"
 9664 . 10245
 /gene="asm29"
 9664 . 10245
 /gene="asm29"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator"
 /protein_id="AA054107.1"
 /db_xref="GI:21449375"
 /transl_table="MSAARBDLVAAALRYMKRQDIAATRAICADMPGPAFLY
 CERSLQELTYLTALTLGLNLSLDAWPAISPDLPVYNVPMFLAAPTARADMPDEAFY
 LGEVLYIAROREPELRELPAMQVREYAKAVVAHLDRLAEAGFDYSDRALLAELVSV
 LDGVAWSMHSRHDGELAKSLDQFAVFAALTR" 10393 . 10857
 /gene="asm28"
 10393 . 10857
 /gene="asm28"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AA054106.1"
 /db_xref="GI:21449374"
 /transl_table="MTDTTTRHRTPYONALALVAGAVELLVGIAGFVGLTNTYDGLFEF
 AGVHSDALALGFVKFKAIVLHFLAFGAGIGIAGLSKAPGARAFILGGGAIYAVLML
 YGVVDEHSTANFVEFLPNAADNMLHLGLAGVMIGLIGALGGRNAHYRAARSD"
 complement(10912 . 11937)
 /gene="asm27"
 complement(10912 . 11937)
 /gene="asm27"
 /codon_start=1
 /transl_table=11
 /product="glucose 1-dehydrogenase"
 /protein_id="AA054105.1"
 /db_xref="GI:21449373"
 /transl_table="NPRRAAEQVNVVVGASSGIGRTTALRLFAAGARVYCAARNTRAL
 DGLVEVKGAGRAVAVTDADIDELAVRVAADVAVRFGGVDTWNVNAGIGVGRVED
 IPAGEFDKRMVRNTYIGHVAGAAALPALRRAGGVYLIGASVYLGLNSAPLQAPVAAK

gene
CDS

```

/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54107.1"
/db_xref="GI:21449375"
/translation="MSAAVRRREOLVAAALRYMKRKGDTAAATTAATCAEADMPHCAPHY
CERSQOEIYTALEITLTGLNTSLDAAMPAPISPLPVDVNTYTLAAFTWAFQADPEAEIV
LGEITVTLQREPELEIRELPAMQHREYVAAVAHDLRLAEAGFDYSVDRLALLVYFSS
LDGLVAASVSHRDGEIAKASIDQFAVMEALTR"
10393..10857
/gene="asm28"
/gene="asm28"
10393..10857

```

```

/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM5106.1"
/db_xref="GI:21449374"
/translation="MTDTTTRHRRPVONALAVAGVELLVGIAQVPELTNTYDGEFF
AGHSVDALAEVFKVSAHLNVHLAFGAAGTAAQLSKAPRAAFLLIGGAIYVIML
GVVDESTANFVFLNADONMLHIGLAVGVIGGALGRNAHYRAARSDD"
complement(10912..11937)
/gene="asm27"
complement(10912..11937)

```

```

/ gene="asm27"
/ codon_start=1
/ transl_table=11
/ product="glucose 1-dehydrogenase"
/ protein_id="AAM54105.1"
/ db_xref="GI:21449373"
/ translation="MPRLAEQVVVVVGASSIGRTVALRPAAGARVVCQAARTRAL
DGLVEVHGAGGAGRAVATADIADEAARAVADVAREGRDITVNAAGICGYRVED
TPAGEFDKVMVNTLIGHVGAKAALPALRRAGSVLLGVASVLKLSAPLQAPYAASK

```

	Query Match	43.2%	Score 43.6	DB 1	Length 82746
	Best Local Similarity	67.8%	Pred. No. 0.43		
	Matches 61	Conservative 0	Matches 29	Indels 0	Gaps
OY	12 GTGCGGCCCTCCCGAAGGCACGTGCAGCTGGATGGCGCGCTCACGACTGGGAATG	71			
Db	14172 GTGCGGCCCTCCCGAGAGCGCGCCGCTGAGACGGCGCGCGCTTCGGGAACTTGC	14113			
OY	72 GACTCGCATGGGGGCTAGACTGGCGAAC	101			
Db	14112 GACTCGATGACGGCGCTGAGACTGGCGAAC	14083			

Search completed: November 6, 2002, 13:32:42
Job time : 606.625 secs

Db 100 CCGGTGGGGGTGTCTATCTGGCCGCGTACCGCGAGCGCCTGCTGGAGAACTCG 159
Oy 69 ACCCCCGCGCGGTTCGCCGCGGTATGCGGCC 100
Db 160 ACCGAAGCCGACTTCGGCGGTGTGCGGCC 191

RESULT 4
LOCUS BM640213 668 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687678931 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
VERSION 19600449622135 5', mRNA sequence.
ACCESSION BM640213
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 668)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: N001004B12 row: M column: 13
Seq primer: M13 Reverse.
Location/Qualifiers
1. 668
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449622135"
/clone_11b="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

BASE COUNT 115 a 196 c 253 g 104 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 668;
Best Local Similarity 60.9%; Pred. No. 18;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 9 CCGCTCCGCGCGGTTCATCGCGCGGTATCTGAGACGCGGCTGCTGATGAGCAA 68
Db 366 CCGGTGGCGCGGTGTCTATCTGGCCGCGGTACGTGCGAGCGCCTGCTGGAGAACTCG 425
Oy 69 ACCCCCGCGCGGTTCGCCGCGGTATGCGGCC 100
Db 426 ACCGAAGCCGACTTCGGCGGTGTGCGGCC 457

RESULT 5
LOCUS BM630368 670 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687501332 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
VERSION 19600449618253 5', mRNA sequence.
ACCESSION BM630368
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 670)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: N001004N8G row: K column: 19
Seq primer: M13 Reverse.
Location/Qualifiers
1. 670
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449618253"
/clone_11b="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

BASE COUNT 119 a 192 c 248 g 111 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 670;
Best Local Similarity 60.9%; Pred. No. 18;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 9 CCGCTCCGCGCGGTTCATCGCGCGGTATCTGAGACGCGGCTGCTGATGAGCAA 68
Db 455 CCGGTGGCGCGGTGTCTATCTGGCCGCGGTACGTGCGAGCGCCTGCTGGAGAACTCG 514
Oy 69 ACCCCCGCGCGGTTCGCCGCGGTATGCGGCC 100
Db 515 ACCGAAGCCGACTTCGGCGGTGTGCGGCC 546

RESULT 6
LOCUS BM606844 694 bp mRNA linear EST 25-FEB-2002
DEFINITION 17000687084364 A.Gam.ad.cDNA blood1 Anopheles gambiae cDNA clone
VERSION 19600449697069 5', mRNA sequence.
ACCESSION BM606844
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 694)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151

KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 670)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: N001004N8G row: K column: 19
Seq primer: M13 Reverse.
Location/Qualifiers
1. 670
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449618253"
/clone_11b="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

BASE COUNT 119 a 192 c 248 g 111 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 670;
Best Local Similarity 60.9%; Pred. No. 18;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Oy 9 CCGCTCCGCGCGGTTCATCGCGCGGTATCTGAGACGCGGCTGCTGATGAGCAA 68
Db 455 CCGGTGGCGCGGTGTCTATCTGGCCGCGGTACGTGCGAGCGCCTGCTGGAGAACTCG 514
Oy 69 ACCCCCGCGCGGTTCGCCGCGGTATGCGGCC 100
Db 515 ACCGAAGCCGACTTCGGCGGTGTGCGGCC 546

RESULT 6
LOCUS BM606844 694 bp mRNA linear EST 25-FEB-2002
DEFINITION 17000687084364 A.Gam.ad.cDNA blood1 Anopheles gambiae cDNA clone
VERSION 19600449697069 5', mRNA sequence.
ACCESSION BM606844
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 694)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151

FEATURES	source
LOCUS	475 bp mRNA linear EST 22-OCT-2001
DEFINITION	BS366493.1
ACCESSION	BS366493
VERSION	1
KEYWORDS	EST.
SOURCE	Hordeum vulgare.
ORGANISM	Hordeum vulgare. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; 1 (bases 1 to 475)
REFERENCE	Wing, R., Close, T.J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Close, S.J., Oates, R. and Malin, D.
AUTHORS	Development of a genetically and physically anchored EST resource for barley genomics. Morex 20 DAP spike cDNA library
TITLE	Unpublished (2001)
JOURNAL	Contact: Wing RA
COMMENT	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Total bg bases = 255 Seq primer: AATTAAACCTCATTAAAGG High quality sequence stop: 405. Location/Qualifiers 1..475

```

/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSME10007D21f"
/clone_lib="Hordeum vulgare 20 DAP spike EST library
HVCNDA0010 (20 DAP)"
/tissue_type="20 DAP spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI;
Plants were grown in the greenhouse at the university of
California, Riverside (Fenton, SJ Close, TJ Close). Whole
spikes with awns trimmed were collected at 20 DAP (Fenton
). Total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 1 million
pfu were in vivo excised to give plasmidscript 3K(-) cDNA
phagemids in the TJ Close lab at the university of
California, Riverside (Choi). Phagemids were plated and
picked at the Clemson University Genetics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders
Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/99pages/bgm/31/clover.html)"

```

Query Match	33.7%	Score 34	DB 12	Length 475
Best Local Similarity	61.1%	Pred. No	23	
Matches	55	Conservative	0	Mismatches 35; Indels 0; Gaps 0;
QY	11	GCTCCGCGCGCTCGTTCATCGCGGCCGATCTCGACACAGGGCGTCTGATGACCAAC	70	
Db	278	GCACGGCCACAGCGCTGTCTGGGAAGGGGAATCTCTCCGACGAGCGGCCCTGTGATGGGCGCAC	219	

QY	71	CCCCGGCGGTTCCGCGGTCATGGCC	100
Db	218	CGGGAGCGGAACCCACGGTCACGGCTC	189

DEFINITION	LOCUS	RESULTS
Tetradodon nigroviridis genome survey sequence T7 end of clone	CNS0224N	697 bp DNA linear GSS 12-MAY-2000

ACCESSION	AL177728
VERSION	AL177728.1
KEYWORDS	GI:7815785
SOURCE	GSS: genome survey sequence, Tetradon nigrovittatus.

REFERENCE
1 (bases 1 to 697)

TITLE
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
JOURNAL
Unpublished

MOORE, D. C., J. L. GILBERT, and J. D. DUNN. 1979. Characterization of the compact genome of the freshwater pufferfish *Tetraodon lineatus*. *Science* 205:1563-1565.

AUTHORING
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES	location/qualifiers
source	1. .697

```

/clone_lib="g"
/note="Genoscope sequence ID : C06g227AB10LP1-end : T7"
BASE COUNT      125 a      215 c      232 g      124 t
ORIGIN           1 others

```

Query Match	33.7%	Score 34	DB 17	Length 697
Best Local Similarity	61.1%	Pred. No. 23		
Matches 55, Conservative	0	Mismatches 35	Indels 0	Gaps 0

OY	7	TGCGCCTCCGGCGGCCTCGTTCATGCGGCCGGGTATCCTTGAGACGGGGCTGCTGATGCAGC	66
Dd	407	TGCTCTCTGCAACGAGCTGCCACGGCACCTCCAAAGTGTACGACGGGCTGTGAAGCAGC	348
OY	67	AAACCCCGCGCGGATTCCGCGGGGTGCATGG	96
Dd	347	AGCTGGCGAGGAGTTGCGCGCCCCCAGG	318

RESULT 11						
CNS042R1/c						
LOCUS	CNS042R1	846 bp	DNA	linear	GSS 18-MAY-2000	
DEFINITION	Tetradon nigriviridis genome survey sequence T7 end of clone 078003 of library G from Tetradon nigriviridis, genomic survey sequence.					
ACCESSION	AL271846					

VERSION	AL271846.1	GI:7993865
KEYWORDS	GSS: genome survey sequence.	
SOURCE	Tetraodon nigroviridis.	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	
REFERENCE	1 (bases 1 to 846)	
AUTHORS	Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissensbach,J.	
TITLE	2 (bases 1 to 846)	
JOURNAL	Tetraodon nigroviridis DNA sequence	
REFERENCE	Unpublished	
AUTHORS	Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissensbach,J.	
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 846)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-APR-2000)	
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .	
FEATURES	Location/Qualifiers	
Source	1..846	
	/organism="Tetraodon nigroviridis"	
	/db_xref="taxon:99883"	
	/clone="078003"	
	/clone_lib="G"	
	/note="Genoscope sequence ID : COR6078AH02LP1-end : 77"	
BASE COUNT	164 a 253 c 273 g 152 t 4 others	
ORIGIN		
Query Match	33.7%; Score 34; DB 17; Length 846;	
Best Local Similarity	61.1%; Pred. No. 23;	
Matches	55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;	
QY	7 TGCCTCCGCGGCGCTGTCATCGGCGCGGTATCTGTGACGACGCGCTGCTGATGCAC 66	
DB	556 TGTCTCTGTGCACCGATCGACGCGCACCTCCACAGTGTACGACGCGCTGTGAAGCAGC 537	
OY	67 AAACCCCGCGCGGTTCGCGCGGTGATG 96	
DB	536 AGCTGGCCGAGAGATTCGCCGCCCCACG 507	
RESULT 12		
LOCUS	CNS02LTD 980 bp DNA linear GSS 14-MAY-2000	
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 146109 of library G from Tetraodon nigroviridis, genomic survey sequence.	
ACCESSION	AL203242	
VERSION	AL203242.1 GI:7861613	
KEYWORDS	GSS: genome survey sequence.	
SOURCE	Tetraodon nigroviridis.	
ORGANISM	Tetraodon nigroviridis.	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	
REFERENCE	1 (bases 1 to 980)	
AUTHORS	Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissensbach,J.	
	Human gene number estimate provided by genome wide analysis using	

[illegible]

RESULT 2
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent NO. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match          35.6%; Score 36; DB 10; Length 36778;
Best Local Similarity 62.0%; Pred. No. 0.046;
Matches 57; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
OY 1 CGGGGATCCGCGTCGCGGCGGTTCATCGGCGCGGTATCTGAGACGAGCGGCTGTA 60
DB 8592 CCGAGACGCCCTCACCCTGCGCGGTCTCCACACCGCGCGCGCTCGACGAGGCAATCGTGG 8651
OY 61 TGCAGCAAAACCCCGCGCGGTTCGCGCGGTC 92
DB 8652 ACAAGCTGACCGCGCGAGAGTCCGCGCGGCG 8683
```

```
RESULT 3
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1
```

```
Query Match          34.1%; Score 34.4; DB 10; Length 15872;
Best Local Similarity 60.9%; Pred. No. 0.13;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
OY 9 CCGCTCCGCGCGTCGTCATGCGCGCGGTATCTGAGACGAGCGGCGGTGATGACAGCA 68
DB 7577 CCGCTACGCGCGCTCGTCGACACGCGCGGAGTCTTCACGACGAGCGGTGAGAGCGCTC 7636
OY 69 ACCCGCGCGCGGTTCGCGCGGTATGCGCGC 100
DB 7637 ACAACGGAACGGGTGAGCGGCTATCGCGCC 7668
```

```
RESULT 4
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
```

```
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474
```

```
Query Match          31.3%; Score 31.6; DB 10; Length 3209;
Best Local Similarity 58.5%; Pred. No. 0.72;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
OY 8 GCCGCTCCGCGCGGTCTTCATGCGCGCGGTATCTGAGACGAGGCTGCTGATGCAGCA 67
DB 724 GCCCGTGGGCGGCGGTCTTCAACCTGCGGTCTTGAGAGATGCGTCTGGAAGAACCA 783
OY 68 AACCCCGCGCGGTTCGCGCGGTATGCGCGCC 101
DB 784 GACCCAGAGTCTTCCAGAGCTCTGCAAGCCC 817
```

```
RESULT 5
US-09-861-289-42
; Sequence 42, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-42
```

```
Query Match          30.9%; Score 31.2; DB 10; Length 846;
Best Local Similarity 57.0%; Pred. No. 0.88;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```

```
OY 1 CGGGGATCCGCGCTCCGCGCGGTTCATGCGCGCGGTATCTGAGACGAGGCGCTGTA 60
DB 446 CCGAGATCCGCGCGGTCTCAGACGAGCGGCTTCCTCCAGAGAGAGCTGCTGCG 505
OY 61 TGCAGCAAAACCCCGCGCGGTTCGCGCGGTATGCGCGC 100
DB 506 GGCTGTGCTGCCCGCGCTCGCAGGAGCTACAAAGCGCGC 545
```

```
RESULT 6
US-09-861-289-3
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
```

APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438051
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: fastseq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 13613
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-3

Query Match 30.9%; Score 31.2; DB 10; Length 13613;
Best Local Similarity 57.0%; Pred. No. 1;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 1 CGGGATCGCCCTCCGGCGCTTCATCGCGCGGTACTGTGACGACGGGCTGTGA 60
DB 409 CCGAGATCCGGCGGCTCAGCGGACCGAGAGCGGTTCTCCAGAGCAGAGCTGTG 468
OY 61 TGCAGCAAAACCCCGCGGCTCCCGCGGCTCATGCGCGC 100
DB 469 GCGTGTGCTGCCCGCGGCTCGCAGCGACTACAAAGCGCGC 508

RESULT 7

US-09-815-242-4083
Sequence 4083, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 4083
LENGTH: 1008
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4083

Query Match 28.3%; Score 28.6; DB 10; Length 1008;
Best Local Similarity 64.2%; Pred. No. 4.8;

Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 31 CGGCGGTATTCCTCGAGCAGCGGCTCTCATGACGAAACCCCGCGGTTCCGCGG 90
DB 521 CGGCCAGTTTCATGCGCGAGTGTGTCTGTGACGATGTGACGAGTATTCGCGCG 580
OY 91 TCATGCG 97
DB 581 GCATGCG 587

RESULT 8

US-09-921-992-2
Sequence 2, Application US/09921992
Patent No. US20020069426A1
GENERAL INFORMATION:
APPLICANT: Boronati, Albert;
APPLICANT: Campos, Narciso;
APPLICANT: Rodriguez-Concepcion, Manuel;
APPLICANT: Rohmer, Michel;
APPLICANT: Seeman, Myriam;
APPLICANT: Valentin, Henry E.;
APPLICANT: Venkatesh, Tyamagondlu V.;
APPLICANT: Venkatesh, Mylavaram;
TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
FILE REFERENCE: 16516, 107/35-21(51897)US
CURRENT APPLICATION NUMBER: US/09/921,992
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/223,483
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 85
SEQ ID NO 2
LENGTH: 33675
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: CDS
LOCATION: (6924)..(7019), (7163)..(7269), (7344)..(7444), (7525)..(7634),
LOCATION: (7694)..(7813), (7923)..(8153), (8253)..(8369), (8515)..(8589),
LOCATION: (9012)..(9071), (9163)..(9223), (9328)..(9472), (9589)..(9730),
LOCATION: (9951)..(10028), (10134)..(10293), (10694)..(10798),
LOCATION: (11028)..(11129)
NAME/KEY: unsure
LOCATION: (1..33675)
OTHER INFORMATION: unsure at all n locations
US-09-921-992-2

Query Match 28.3%; Score 28.6; DB 10; Length 33675;
Best Local Similarity 61.3%; Pred. No. 5.6;
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 27 CATGCGGCGGTATTCCTGAGCAGCGGCTGTGATGACGAAACCCCGCGGTTCCGC 86
DB 5429 CAGCAGCGCATGCGCCAGCGGAGCGGCTGTGACACGAGTCCACGACGCGCGC 5488
OY 87 GCGGTATGCGCGCC 101
DB 5489 GCGGTGCGGCGCGC 5503

RESULT 9

US-09-815-242-7845
Sequence 7845, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

```
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7845
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1506)
US-09-815-242-7845
```

```
Query Match          28.1%; Score 28.4; DB 10; Length 1506;
Best Local Similarity 58.1%; Pred. No. 5.6;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
QY 13 TCCGCGCGCTTCATCGCGCGGTCTCTGACGACGCGGCTGCTATGACGAAACC 72
    |||||  |||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 569 TCCGCGCTTCCTCAGCGCGGCTTCCTCGAAGTGAGACCCCGCTGCGAGACA 628
QY 73 CCGCGCGGTTCGCGCGGTATGCGC 98
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 629 TCCCGCGCGCGCGCGCGCGCAAGCCG 654
```

```
RESULT 10
US-09-867-550-303
; Sequence 303, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Foad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1a Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 303
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(2)
; OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-303
```

```
Query Match          27.7%; Score 28; DB 10; Length 246;
```

```
Best Local Similarity 56.5%; Pred. No. 6.7;
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 9 CCGCTCCGCGCGCTTCATGCGCGGTATCTGAGACGCGGCTGCTGATGACGCA 68
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 40 CCGACCGCGATGTCGCGCGCGGCGGACATCTGCGCGCATGCGCGTGGAGATGCGC 99
QY 69 ACCCGCGCGGTTCGCGCGGTATGCGCGC 100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 100 ACCGAGACCGCTTCGCGGACGCGATCTCTC 131
```

```
RESULT 11
US-09-815-242-7789
; Sequence 7789, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Traawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7789
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1032)
US-09-815-242-7789
```

```
Query Match          27.5%; Score 27.8; DB 10; Length 1032;
Best Local Similarity 55.8%; Pred. No. 8.1;
Matches 53; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 5 GATGCGCTTCGCGCGGTCTATGCGCGGTATCTGAGACGCGGCTGCTGATGCA 64
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 432 GCTGCTGCTGCTGCGGCTGCGGATCGCGCGCGGCGGCGGCGGCGGCTGCTGCT 491
QY 65 GCAAAACCGCGCGGTTCGCGCGGTATGCGCGC 99
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 492 GCGCGGGGTGCGGCTGCTGCTGCTGCTGCGC 526
```

```
RESULT 12
US-10-002-600-62
; Sequence 62, Application US/10002600
; Patent No. US2002013707A1
; GENERAL INFORMATION:
```


APPLICANT: Hopkins, Christopher M.
APPLICANT: Peterson, David P.
APPLICANT: Cocks, Benjamin G.
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
FILE REFERENCE: PA-0042 US
CURRENT APPLICATION NUMBER: US/10/002,600
CURRENT FILING DATE: 2001-10-25
PRIORITY APPLICATION NUMBER: 60/243,521
PRIORITY FILING DATE: 2000-10-25
NUMBER OF SEQ ID NOS: 116
SOFTWARE: PERL Program
SEQ ID NO 62
LENGTH: 1699
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Template ID: 23849.1
US-10-002-600-62

Query Match 27.3%; Score 27.6; DB 12; Length 1699;
Best Local Similarity 60.8%; Pred. No. 9.4;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 4 GGATGCCGCTCCGGCGCTGTCATGCGCGGATCTGGAGAGCGGCTGATGC 63
DB 153 GGGTGGCTTCCGGACCGGACATGACGCGGCCACTAAGACGAGCTGGACCC 212

OY 64 AGCAACCCCGCGCG 77
DB 213 TACAACCTGGCGCG 226

RESULT 13
US-09-925-301-441/C
Sequence 441, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIORITY APPLICATION NUMBER: PCT/US00/05882
PRIORITY FILING DATE: 2000-03-08
PRIORITY APPLICATION NUMBER: 60/124,270
PRIORITY FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 441
LENGTH: 1082
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1136)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (462)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (465)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1074)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-441

Query Match 26.9%; Score 27.2; DB 10; Length 1082;
Best Local Similarity 55.2%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 4 GGATGCCGCTCCGGCGCTGTCATGCGCGGATCTGGAGAGCGGCTGATGC 63
DB 257 GGAGCAGACGGCGCGCGGCTGTGGTCCCGCAGAGGACAGCCGCTCTGATGC 198
OY 64 AGCAACCCCGCGCGGTTCCGGGTCATGCGCG 99
DB 197 GCCACAGTCCCGGAGCTCCGAGCTCGCGGCGCTC 162

RESULT 14
US-09-768-826-27
Sequence 27, Application US/09768826
Patent No. US20020012966A1
GENERAL INFORMATION:
APPLICANT: Shi et al.
TITLE OF INVENTION: 18 human secreted proteins
FILE REFERENCE: PF512P1
CURRENT APPLICATION NUMBER: US/09/768,826
CURRENT FILING DATE: 2001-01-25
PRIORITY APPLICATION NUMBER: PCT/US00/22350
PRIORITY FILING DATE: 2000-08-15
PRIORITY APPLICATION NUMBER: 60/148,759
PRIORITY FILING DATE: 1999-08-16
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 27
LENGTH: 1558
TYPE: DNA
ORGANISM: Homo sapiens
US-09-768-826-27

Query Match 26.7%; Score 27; DB 10; Length 1558;
Best Local Similarity 56.0%; Pred. No. 14;
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 4 GGATGCCGCTCCGGCGCTGTCATGCGCGGATCTGGAGAGCGGCTGATGC 63
DB 39 GGAGACCGCGCGCGGACGAGAGTGTGCGGGGGAGACTGTGCTCTGATC 98

OY 64 AGCAACCCCGCGCGGTTCCGGGTCATG 94
DB 99 CGCCCCCACCCTCTCTTGTGACTGCGGT 129

RESULT 15
US-09-944-413-77
Sequence 77, Application US/09944413
Patent No. US20020156004A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Baton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hlavin, Kenneth
APPLICANT: Kijavitt, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P254BP1C1
CURRENT APPLICATION NUMBER: US/09/944,413
CURRENT FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: 09/866,028
PRIORITY FILING DATE: 2001-05-25
PRIORITY APPLICATION NUMBER: 60/067,411

PRIOR FILING DATE:	December 3, 1997
PRIOR APPLICATION NUMBER:	60/069,334
PRIOR FILING DATE:	December 11, 1997
PRIOR APPLICATION NUMBER:	60/069335
PRIOR FILING DATE:	December 11, 1997
PRIOR APPLICATION NUMBER:	60/069,278
PRIOR FILING DATE:	December 11, 1997
PRIOR APPLICATION NUMBER:	60/069,425
PRIOR FILING DATE:	December 12, 1997
PRIOR APPLICATION NUMBER:	60/069,696
PRIOR FILING DATE:	December 16, 1997
PRIOR APPLICATION NUMBER:	60/069,694
PRIOR FILING DATE:	December 16, 1997
PRIOR APPLICATION NUMBER:	60/069,702
PRIOR FILING DATE:	December 16, 1997
PRIOR APPLICATION NUMBER:	60/069,870
PRIOR FILING DATE:	December 17, 1997
PRIOR APPLICATION NUMBER:	60/069,873
PRIOR FILING DATE:	December 17, 1997
PRIOR APPLICATION NUMBER:	60/068,017
PRIOR FILING DATE:	December 18, 1997
PRIOR APPLICATION NUMBER:	60/070,440
PRIOR FILING DATE:	January 5, 1998
PRIOR APPLICATION NUMBER:	60/074,086
PRIOR FILING DATE:	February 9, 1998
PRIOR APPLICATION NUMBER:	60/074,092
PRIOR FILING DATE:	February 9, 1998
PRIOR APPLICATION NUMBER:	60/075,945
PRIOR FILING DATE:	February 25, 1998
PRIOR APPLICATION NUMBER:	60/112,850
PRIOR FILING DATE:	December 16, 1998
PRIOR APPLICATION NUMBER:	60/113,296
PRIOR FILING DATE:	December 22, 1998
PRIOR APPLICATION NUMBER:	60/146,222
PRIOR FILING DATE:	July 28, 1999
PRIOR APPLICATION NUMBER:	PCT/US98/19330
PRIOR FILING DATE:	September 16, 1998
PRIOR APPLICATION NUMBER:	PCT/US98/25108
PRIOR FILING DATE:	December 1, 1998
PRIOR APPLICATION NUMBER:	09/216,021
PRIOR FILING DATE:	December 16, 1998
PRIOR APPLICATION NUMBER:	09/218,517
PRIOR FILING DATE:	December 22, 1998
PRIOR APPLICATION NUMBER:	09/234,311
PRIOR FILING DATE:	March 3, 1999
PRIOR APPLICATION NUMBER:	PCT/US99/12252
PRIOR FILING DATE:	June 22, 1999
PRIOR APPLICATION NUMBER:	PCT/US99/21090
PRIOR FILING DATE:	September 15, 1999
PRIOR APPLICATION NUMBER:	PCT/US99/28409
PRIOR FILING DATE:	No. US20020156004A1ember 30, 1999
PRIOR APPLICATION NUMBER:	PCT/US99/28313
PRIOR FILING DATE:	No. US20020156004A1ember 30, 1999
PRIOR APPLICATION NUMBER:	PCT/US99/28301
PRIOR FILING DATE:	December 1, 1999
PRIOR APPLICATION NUMBER:	PCT/US99/30095
PRIOR FILING DATE:	December 16, 1999
PRIOR APPLICATION NUMBER:	PCT/US00/03565
PRIOR FILING DATE:	February 11, 2000
PRIOR APPLICATION NUMBER:	PCT/US00/04414
PRIOR FILING DATE:	February 22, 2000
PRIOR APPLICATION NUMBER:	PCT/US00/05841
PRIOR FILING DATE:	March 2, 2000
PRIOR APPLICATION NUMBER:	PCT/US00/08439
PRIOR FILING DATE:	March 30, 2000
PRIOR APPLICATION NUMBER:	PCT/US00/14042
PRIOR FILING DATE:	May 22, 2000
PRIOR APPLICATION NUMBER:	PCT/US00/20710
PRIOR FILING DATE:	July 28, 2000
PRIOR APPLICATION NUMBER:	PCT/US00/32678
PRIOR FILING DATE:	December 1, 2000
PRIOR APPLICATION NUMBER:	PCT/US01/06520
PRIOR FILING DATE:	February 28, 2001

```

; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 77
; LENGTH: 2849
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-944-413-77

```

Query Match	26.7%	Score 27	DB 9	Length 2849
Best Local Similarity	56.0%	Pred. No. 14		
Matches 51; Conservative	0	Mismatches 40	Indels 0	Gaps 0

QY 4 GGAGTCCGCTCCGGGGCTGTTATGGGGCGGTATCCGAGACGGGCTGCTGATGC 63
 Db 66 GGAGACCGCGCCGGGGAGACGAGAGTGTGTTGGGTGGGGGGAGCCTTGCGTCTCGTAC 126
 QY 64 AGCAAAACCCCGCGCGGTCTCCGGCGCGGTCAAT 94
 Db 126 CGCCCCCACCCTCTCTTGTGCACTGCGGT 156

```
Search completed: November 6, 2002, 20:33:11
Job time : 46.125 secs
```

3

; Sequence 1, Application US/0

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 42652
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 42652
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 42652
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42552 CGGGATGCCGCTCCGCGGCTGTCATGCGCGGATCTGAGACGAGGCGCTGTA 42611
```

```
QY 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGAGCGCC 42652
```

RESULT 6

US-09-568-486-1

; Sequence 1, Application US/09568486
; Patent No. 635459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match

Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 7

US-09-568-472-1

; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match

Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8

US-09-567-899-1

; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match

Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCCCCCTCCGCGGCTGCTCATGCGCGGATCTGAGACAGCGGCTGCTGA 60
DB 42552 CGGGATGCCCCCTCCGCGGCTGCTCATGCGCGGATCTGAGACAGCGGCTGCTGA 42611
QY 61 TGCAGCAAAACCCCGCGGCTTCCGCGGATGAGCCGCC 101

DB 42612 TGCAGCAAAACCCCGCGGCTTCCGCGGATGAGCCGCC 42652

RESULT 9

US-09-036-987A-1

; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Balitz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Paul J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998

Query Match

Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATGCCCCCTCCGCGGCTGCTCATGCGCGGATCTGAGACAGCGGCTGCTGA 60
DB 42552 CGGGATGCCCCCTCCGCGGCTGCTCATGCGCGGATCTGAGACAGCGGCTGCTGA 42611

QY 61 TGCAGCAAAACCCCGCGGCTTCCGCGGATGAGCCGCC 101
DB 42612 TGCAGCAAAACCCCGCGGCTTCCGCGGATGAGCCGCC 42652

RESULT 9

US-09-036-987A-1

; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Balitz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Paul J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 44.6%; Score 45; DB 3; Length 80161;
Best Local Similarity 65.3%; Pred. No. 0.00047;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGGCTTCATGCGCGGATCTGAGACGAGCGGCTGCTGA 60
DB 74831 CGGAAACCGCTTCCGGGTGTGTGACGCCCGCGGTGTGTGATGACGCTGTCTGA 74890

QY 61 TGCAGCAACCCCGCGCGGCTTCGCGCGGTCATGCGGCC 101
DB 74891 TGTGATGTCCCGGAGCGCTTGACGCGGCTGTGGCGCC 74931

RESULT 10
US-09-370-700-1
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patl J
APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 44.6%; Score 45; DB 4; Length 80161;
Best Local Similarity 65.3%; Pred. No. 0.00047;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGGCTTCATGCGCGGATCTGAGACGAGCGGCTGCTGA 60
DB 74831 CGGAAACCGCTTCCGGGTGTGTGACGCCCGCGGTGTGTGATGACGCTGTCTGA 74890

QY 61 TGCAGCAACCCCGCGCGGCTTCGCGCGGTCATGCGGCC 101
DB 74891 TGTGATGTCCCGGAGCGCTTGACGCGGCTGTGGCGCC 74931

RESULT 11
US-08-258-261B-6
Sequence 6, Application US/08258261B
Patent No. 5639949

GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

Query Match 39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0099;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CGGCTCCGGCGGCTGTTCATGCGCGGATCTGAGACGAGCGGCTGTATGACGCA 68
DB 2492 CCGCTACGCGCGCTGTGACGCGCGGCGGCGGCTGTGACGATGGCTGATGACGACANG 2551
QY 69 ACCCGCGGCGGCTTCGCGCGGTCATGCGGCC 101
DB 2552 AGCCCGAGGCGATGACGCGGCTTTGCTCCC 2584

RESULT 12
US-08-456-837-6
Sequence 6, Application US/08456837
Patent No. 5643774
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew

```

: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/456,837
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEITICAL: NO
: ANTI-SENSE: NO
: US-08-456-837-6

Query Match          39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CCGCTCCCGCGGCGTTCATGCGCGCGGTATCTGAGACGAGCGGCTGCTGATGCAGCAA 68
DB 2492 CCGCTCAGCGCGCGTGTGACGCGCGCGCGCTTGACGATGGGCTGATCAGCGCATG 2551
QY 69 ACCCCCGCGGCGTTCGCGCGGCGTCAATGCGGCC 101
DB 2552 AGCCCGAGCGCATGACCGCGCTTTCCTCC 2584

RESULT 13
US-08-457-342-6
: Sequence 6, Application US/08457342
: Patent No. 5662898
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph

```

```

: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,342
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEITICAL: NO
: ANTI-SENSE: NO
: US-08-457-342-6

Query Match          39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CCGCTCCCGCGGCGTTCATGCGCGCGGTATCTGAGACGAGCGGCTGCTGATGCAGCAA 68
DB 2492 CCGCTCAGCGCGCGTGTGACGCGCGCGCGCTTGACGATGGGCTGATCAGCGCATG 2551
QY 69 ACCCCCGCGGCGTTCGCGCGGCGTCAATGCGGCC 101
DB 2552 AGCCCGAGCGCATGACCGCGCTTTCCTCC 2584

RESULT 14
US-08-457-646A-6
: Sequence 6, Application US/08457646A
: Patent No. 5679560
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Ciba-Gelby Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 530
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-457-646A-6

Query Match          39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CCGCTCCGGCGCGCTTCATGCGCGCGGTATCTGGACGACGCGGCTGTGATGACAGCA 68
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2492 CCGCTCAGCGCGCTCGTGCAGCGCGCGCGCGCGCTTGACGATGGCTGATCAGCAGATG 2551
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 69 ACCCGCGCGGTTCCGCGCGGTATGCGCGCC 101
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2552 AGCCCGAGCGCATGCGACGCGCTTTGCTCC 2584
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 15
US-08-458-076A-6
: Sequence 6, Application US/08458076A
: Patent No. 5698425
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Gelby Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
```

```

: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,076A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-458-076A-6

Query Match          39.8%; Score 40.2; DB 1; Length 28958;
Best Local Similarity 64.5%; Pred. No. 0.0089;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 9 CCGCTCCGGCGCGCTTCATGCGCGCGGTATCTGGACGACGCGGCTGTGATGACAGCA 68
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2492 CCGCTCAGCGCGCTCGTGCAGCGCGCGCGCGCGCTTGACGATGGCTGATCAGCAGATG 2551
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 69 ACCCGCGCGGTTCCGCGCGGTATGCGCGCC 101
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2552 AGCCCGAGCGCATGCGACGCGCTTTGCTCC 2584
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: November 6, 2002, 17:29:13
Job time : 211.625 secs

PR 13-APR-2001; 2001US-269020P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosla C;
PI Lau J, Licardl PJ, Regentin R, Santl D, Tang L;
XX
XX WPI: 2002-075167/10.
DR P-PSDB: ABB07167.
XX
PT Recombinant host cells useful for producing polyketides e.g. epothilone
PT or its derivatives, comprises a recombinant expression vector encoding
PT a heterologous polyketide synthase gene -
PS
PS Example 11; Page 164-165; 221pp: English.
XX
XX The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of the
CC KR domain of extender module 6 of the epothilone PKS. Inactivation of
CC this domain results in a novel PKS capable of producing a 9-keto-
CC epothilone analogue.
XX
SQ Sequence 552 BP; 70 A; 176 C; 214 G; 92 T; 0 other;
Query Match 100.0%; Score 101; DB 24; Length 552;
Best Local Similarity 100.0%; Pred. No. 6.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGGGATGCCGTCGCCGCGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 60
DB 236 CGGGGATGCCGTCGCCGCGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 295
QY 61 TGCAGCAAAACCCCGCGCGGTTCCGCGGCTCATGCGGCC 101
DB 296 TGCAGCAAAACCCCGCGCGGTTCCGCGGCTCATGCGGCC 336
RESULT 2
ID AA168064 standard; DNA; 552 BP.
XX
XX AA168064:
AC
XX
XX 13-MAR-2002 (first entry)
XX
XX 9-keto-epothilone PKS inactive KR domain nucleotide sequence.
DE
XX Cystobacterineae; recombinant; polyketide synthase; PKS; polyketide;
KW epothilone; p11a gene; cytosatic; antiproliferic; antiarthritic;
KW antilactosclerotic; antiinflammatory; neuroprotective; vasotropic; ds.
XX
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..552
FT /tag= a
FT /note= "mutated KR domain of extender module 6"
XX
XX WO200183800-A2.
XX
XX

PD 08-NOV-2001.
XX
XX
PF 26-APR-2001; 2001WO-US13793.
XX
XX 28-APR-2000; 2000US-0560367.
PR 14-SEP-2000; 2000US-232696P.
PR 21-DEC-2000; 2000US-257517P.
PR 03-APR-2001; 2001US-0825856.
PR 03-APR-2001; 2001US-0825876.
PR 13-APR-2001; 2001US-269020P.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosla C;
PI Lau J, Licardl PJ, Regentin R, Santl D, Tang L;
XX
XX WPI: 2002-075167/10.
DR P-PSDB: ABB07168.
XX
XX Recombinant host cells useful for producing polyketides e.g. epothilone
PT or its derivatives, comprises a recombinant expression vector encoding
PT a heterologous polyketide synthase gene -
PS
PS Example 11; Page 165-166; 221pp: English.
XX
XX The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of a
CC mutated and inactive KR domain of extender module 6 of the novel 9-keto-
CC epothilone PKS of the present invention.
XX
SQ Sequence 552 BP; 70 A; 177 C; 212 G; 93 T; 0 other;
Query Match 100.0%; Score 101; DB 24; Length 552;
Best Local Similarity 100.0%; Pred. No. 6.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGGGGATGCCGTCGCCGCGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 60
DB 236 CGGGGATGCCGTCGCCGCGCTTCATCGCGCGGTATCTGACGACGCGGCTCTGA 295
QY 61 TGCAGCAAAACCCCGCGCGGTTCCGCGGCTCATGCGGCC 101
DB 296 TGCAGCAAAACCCCGCGCGGTTCCGCGGCTCATGCGGCC 336
RESULT 3
ID AAA29349 standard; DNA; 71989 BP.
XX
XX AAA29349:
AC
XX
XX 12-SEP-2000 (first entry)
XX
XX Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
DE
XX Epothilone; polyketide synthase; epOA; epOB; epOC; epOD; epOE; epOF;
KW epOL; epOK; P450 epoxidase; ORFV; ORFP; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytosatic; ds.
XX
XX Sorangium cellulosum.
OS

```
XX Key Location/Qualifiers
FH CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /tag= c
FT /label= epOA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading domain"
FT misc_RNA 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading domain"
FT misc_RNA 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading domain, potentially involved in formation of the thiazole moiety"
FT misc_RNA 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading domain"
FT CDS 6260..10493
FT /tag= h
FT /label= epOB_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS module"
FT misc_RNA 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS module"
FT misc_RNA 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS module"
FT misc_RNA 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS module"
FT CDS 10639..16137
FT /tag= x
FT /label= epOC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epOD_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
```

```
FT misc_RNA 29433..30287
FT /tag= ap
FT /note= "encodes ER5"
FT misc_RNA 30321..30869
FT /tag= ag
FT /note= "encodes KR5"
FT misc_RNA 31077..31373
FT /tag= ar
FT /note= "encodes ACP5"
FT misc_RNA 31440..32807
FT /tag= as
FT /note= "encodes K56"
FT misc_RNA 33018..34067
FT /tag= at
FT /note= "encodes AAT6"
FT misc_RNA 34107..34676
FT /tag= au
FT /note= "encodes DH6"
FT misc_RNA 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA 36705..37256
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA 37470..37769
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS 37912..49308
FT /tag= ay
FT /label= epro_gene
FT /note= "encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /tag= az
FT /note= "encodes K57"
FT misc_RNA 39589..40626
FT /tag= ba
FT /note= "encodes AAT7"
FT misc_RNA 41341..41922
FT /tag= db
FT /note= "encodes KR7"
FT misc_RNA 42181..42423
FT /tag=

Query Match 100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 7.5e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATGCGCGCGCGCTCATCGCGCGCGGTATCCTCGACGACGCGGCTGCTGA 60
Db 36940 CGGGATGCGCGCGCGCTCATCGCGCGGTATCCTCGACGACGCGGCTGCTGA 36999
QY 61 TGCAGCAAAACCCCGCGGTTCCGCGCGTCAATGCGGCC 101
Db 37000 TGCAGCAAAACCCCGCGGTTCCGCGCGTCAATGCGGCC 37040

RESULT 4
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX AAZ55887;
XX 10-APR-2000 (first entry)
XX Sorangium cellulosum 68.75 kb contig.
DE Epochlone biosynthesis: type I polyketide synthase; taxol substitute;
KM anticancer; ds.
OS Sorangium cellulosum.
XX Key Location/Qualifiers
FH 1..1826
FT CDS /tag= a
```

```
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT complement (5612..5992)
FT /tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /tag= f
FT /product= "Type I polyketide synthase, EPOS A (AAV58573)"
FT 11872..116104
FT /tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P (AAV58574)"
FT 16251..21749
FT /tag= h
FT /product= "Type I polyketide synthase, EPOS B (AAV58575)"
FT 21746..43519
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C (AAV58576)"
FT 43524..54920
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D (AAV58577)"
FT 54935..62254
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E (AAV58578)"
FT 62369..63628
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F (AAV58579)"
FT 63779..64333
FT /tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT complement (63853..64290)
FT /tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT complement (64287..64727)
FT /tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /tag= q
FT /product= "Orf 10 protein"
FT complement (65008..65874)
FT /tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT complement (65871..66338)
FT /tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT 67334..68251
FT /tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT 68346..68750
FT /partial
FT /tag= v
```

```
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
XX
XX
XX MO9966028-A2.
XX
XX 23-DEC-1999.
XX
XX
XX 16-JUN-1999; 99WO-EP04171.
XX
XX 18-JUN-1998; 98US-0099504.
XX 24-SEP-1998; 98US-0101631.
XX 05-FEB-1999; 99US-0118906.
XX
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERINDUNGEN VERW GES MBH.
XX
XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX
XX WPI: 2000-097741/08.
XX P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
XX AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
XX AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
XX AAY58592, AAY58593, AAY58594.
XX
XX New isolated epothilone synthase genes, used for the recombinant
XX production of epothilone for use in cancer therapy -
XX
XX Claim 14; Page 87-104; 174pp; English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX comprising 22 open reading frames (ORFs) and includes genes encoding
XX proteins involved in the biosynthesis of epothilones. Epothilones A and
XX B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX starter unit; polyketides being synthesised from two-carbon building
XX blocks, the beta-carbon of which always carries a keto group. Each round
XX of two-carbon addition is carried out by a complex of enzymes known as
XX the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
XX the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
XX and EPOS E (AAY58575-58578) are involved in polyketide backbone
XX formation. EPO F (AAY58579) is an epothilone macroactone oxidase, and
XX the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
XX involved in transport. Epothilones mimic the biological activity of
XX taxol, and may be substituted for taxol in cancer chemotherapeutic
XX compositions. Epothilones exhibit a much lower drop in potency against a
XX multiply drug-resistant cell line compared with taxol, and are
XX considerably less efficiently exported from such cells by the multidrug
XX resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX epothilones as anticancer agents, they are problematical to produce on a
XX large scale. Epothilones are too complex for industrial scale chemical
XX synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX poor yields of epothilones. The nucleic acids of the invention may be
XX used for the recombinant production of epothilones in a heterologous host
XX that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX Query Match 98.4%; Score 99.4; DB 21; Length 68750;
XX Best Local Similarity 99.0%; Pred. No. 2e-17;
XX Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 CGGGGATGCCGCTCGGGCGCTTCATGCGCGCGGATCGTGCAGACGAGGCTGTGCA 60
XX |||||
XX Db 42552 CGGGGATGCCGCTCGGGCGCTTCATGCGCGCGGATCGTGCAGACGAGGCTGTGCA 42611
XX |||||
XX Qy 61 TGCAGCAAAACCCCGCGGGGTTCCGCGGCGATGAGGCCCC 101
XX |||||
XX Db 42612 TGCAGCAAAACCCCGCGGGGTTCCGCGGCGATGAGGCCCC 42652
XX |||||
XX
XX RESULT 5
XX AAF90035
XX ID AAF90035 standard; DNA; 8301 BP.
```

```
XX
XX AAF90035;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon: polyketide; polyketide antibiotic;
XX type I polyketide synthase; ss.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 1..8301
XX FT /*tag= a
XX FT /product= "type I polyketide synthase"
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR03311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
XX
XX Jeanin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
XX Cappellano C, Franco F, Raynal A, Ball M, Sezonov G, Tuphile K;
XX Frostegard A;
XX
XX WPI: 2001-374849/39.
XX P-PSDB: AAB83972.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics -
XX
XX Claim 35; Page 302-305; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through an anion exchange
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthesis) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. AAF90034-39
XX represent open reading frames (ORFs) of the coding strand of cosmid
XX a26g1, and encode type I polyketide synthases.
XX
XX Sequence 8301 BP; 1426 A; 2629 C; 2669 G; 1577 T; 0 other;
XX
XX Query Match 46.7%; Score 47.2; DB 22; Length 8301;
XX Best Local Similarity 69.6%; Pred. No. 0.0012;
XX Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
XX
XX 9 CGCGTCCGCGGGGTCATGCGCGCGGATATCCGAGACGAGCGGCTGTGTCAGCA 68
XX |||||
XX Db 7435 CGCGTCCGCGGGGTCATGCGCGCGGATATCCGAGACGAGCGGCTGTGTCAGCA 7494
XX |||||
XX Qy 69 ACCCGCGCGGATTCGCGCGGATGATGCGGCC 100
XX |||||
XX Db 7495 GACGTGACGCGGATGCAAGCGTATGCGGCC 7526
XX |||||
XX
XX RESULT 6
XX AAF90033
XX ID AAF90033 standard; DNA; 34071 BP.
```

XX AAF90033;
AC 06-AUG-2001 (first entry)
XX
DT Nucleotide sequence of cosmid a26g1 (coding strand).
XX
DE Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
KW Synthetic.
XX
OS WO200140497-A2.
XX
PN 07-JUN-2001.
XX
PD 27-NOV-2000; 2000MO-FR03311.
XX
PF 29-NOV-1999; 99FR-0015032.
XX
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tiphlie K;
PI Frostegard A;
XX
DR WPI; 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics .
XX
XX Example 14; Page 289-300; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthetase) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand), which encodes different
CC polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
Query Match 46.7%; Score 47.2; DB 22; Length 34071;
Best Local Similarity 69.6%; Pred. No. 0.0012;
Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 9 CCGCTCCGGCGCTTCATGCGCGCGGTATCTGGACGACGGCTGCTGATGACGAA 68
DB 12067 CCGCTCGCGCGCTTCATGCGCGCGGTGCTGACGACGCGTACTGTAATCAAG 12126
QY 69 ACCCCGCGGGGTTCGCGCGCGGTATGCGGCC 100
DB 12127 GACTGACGCGGATCCCAAGCGTCAATGCGCC 12158

RESULT 7
AAF90032/C
ID AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX

KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS Synthetic.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000MO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
XX
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tiphlie K;
PI Frostegard A;
XX
DR WPI; 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics .
XX
XX Example 14; Page 274-288; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthetase) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC encodes different polyketide synthases.
XX
SQ Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
Query Match 46.7%; Score 47.2; DB 22; Length 42717;
Best Local Similarity 69.6%; Pred. No. 0.0012;
Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 9 CCGCTCCGGCGCTTCATGCGCGCGGTATCTGGACGACGGCTGCTGATGACGAA 68
DB 22053 CCGCTCGCGCGCTTCATGCGCGCGGTGCTGACGACGCGTACTGTAATCAAG 21994
QY 69 ACCCCGCGGGGTTCGCGCGCGGTATGCGGCC 100
DB 21993 GACTGACGCGGATCCCAAGCGTCAATGCGCC 21962

RESULT 8
AAF88339
ID AAF88339 standard; DNA; 16767 BP.
XX
XX AAF88339;
XX
XX 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment encoding ORF22, SEQ ID 49.
XX
XX Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
OS Saccharopolyspora spinosa.
XX
XX DE19957268-A1.

[illegible]

```

XX 27-AUG-1999; 99DE-1040596.
XX
XX (FARB ) BAYER AG.
XX
XX Eberz G, Moehrl V, Froede R, Velden R, Salas JA;
XX
XX WPI: 2001-267102/28.
XX
XX
XX
XX
XX Claim 7; Page 92-102; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactivate or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylrharnnose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine and trimethylrharnnose biosynthesis.
XX
XX
XX
XX
XX Sequence 29736 BP; 4401 A; 10346 C; 10080 G; 4909 T; 0 other;
XX
XX
XX
XX
XX Query Match 44.6%; Score 45; DB 22; Length 29736;
XX Best Local Similarity 65.3%; Pred. No. 0.0047;
XX Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
XX
XX
XX 1 CGGGGATGCGCGCTCGCGCGCTGTCATGACGCGCGGTAFCCTGGACGAGGGCTGCTGA 60
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 3836 CGGAAACCCGCTGGCGGGGTGTGTGCACGCCGCCGGTGTCTGATGACGGTGTCTCA 3895
XX
XX 61 TGCAGCAAAACCCCGCGCGGTTCCGCGCGGTGATGAGCGCC 101
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 3896 TGTGATGTGCGCGGAGCGCTTGACGCGGTTCGCGGCC 3936
XX
XX
XX RESULT 10
XX AAF88313
XX ID AAF88313 standard; DNA; 50000 BP.
XX
XX
XX AAF88313;
XX
XX
XX 28-AUG-2001 (first entry)
XX
XX
XX S. spinosa DNA fragment SEQ ID 2.
XX
XX
XX
XX
XX Forosamine; trimethylrharnnose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
XX
XX
XX Saccharopolyspora spinosa.
XX
XX
XX DE19957268-A1.
XX
XX
XX 08-MAR-2001.
XX
XX
XX 29-NOV-1999; 99DE-1057268.
XX
XX
XX 27-AUG-1999; 99DE-1040596.
XX
XX

```

PA (FARB) BAYER AG.
 XX Eberz G, Moehle V, Froede R, Velten R, Salas JA:
 XX WPI: 2001-267102/28.
 DR
 XX
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 PT
 XX
 PS Claim 7: Page 31-49; 354pp; German.
 XX
 CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (1) to
 CC identify, inactive or modulate genes involved in the biosynthesis of
 CC (II); (11) to generate a library of polypeptide synthetases; (111) for
 CC adding forosamine or trimethylrhamsome to a spinosyn or polypeptide
 CC aglycone; and (IV) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrocides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
 CC genome which contains the coding regions for proteins involved in
 CC forosamine, trimethylrhamsome and polypeptide synthase biosynthesis.
 CC
 XX
 SQ Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 other;
 Query Match 44.6%; Score 45; DB 22; Length 50000;
 Best Local Similarity 65.3%; Pred. No. 0 0048;
 Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 1 CGGGGATGCCGTCGCGGCGGTCGTCATGCGCGGATATCTGACGACGGGCTGCTGA 60
 DB 49460 CGGAAAACCGCTGCGGGTGTGTCACACGCCGCGGTGCTGATGACGTCGTCTGA 49519
 QY 61 TGCAGCAAAACCCCGCGCGGTCGCGGCTCATGCGGCC 101
 DB 49520 TGTGATGTGTCGCGGAGCGCTTGACGCGGCTGCGGCC 49560
 RESULT 11
 AA221501
 ID AA221501 standard; DNA: 80161 BP.
 AC AA221501;
 XX
 DT 01-DEC-1999 (first entry)
 XX
 DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
 XX
 KW spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
 KW microtides; arachnid; nematode; insect; polypeptide; polypeptide synthase;
 KW PKS; extender module; initiator module; acyl transferase domain; AT;
 KW acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
 KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
 KW insecticide; ss.
 XX
 OS Saccharopolyspora spinosa.
 XX
 FH Key Location/Qualifiers
 FT CDS complement (1135..1971)
 FT /tag= a
 FT /product= ORFL16
 FT /note= "Protein involved in transcription control"
 FT 2024..2791
 FT /tag= b
 FT /product= ORFL15

FT /note= "keto acyl reductase"
 FT complement (3416..4165)
 FT /tag= c
 FT /product= spns
 FT /note= "spinosyn biosynthesis protein S"
 FT complement (4168..5325)
 FT /tag= d
 FT /product= spnr
 FT /note= "spinosyn biosynthesis protein R"
 FT complement (5363..6751)
 FT /tag= e
 FT /product= spnq
 FT /note= "spinosyn biosynthesis protein Q"
 FT 7083..8450
 FT /tag= f
 FT /product= spnp
 FT /note= "spinosyn biosynthesis protein P"
 FT 8967..10427
 FT /tag= g
 FT /product= spno
 FT /note= "spinosyn biosynthesis protein O"
 FT 10436..11434
 FT /tag= h
 FT /product= spnn
 FT /note= "spinosyn biosynthesis protein N"
 FT complement (11530..12492)
 FT /tag= i
 FT /product= spnm
 FT /note= "spinosyn biosynthesis protein M"
 FT complement (12696..13547)
 FT /tag= j
 FT /product= spnl
 FT /note= "spinosyn biosynthesis protein L"
 FT complement (13592..14785)
 FT /tag= k
 FT /product= spnk
 FT /note= "spinosyn biosynthesis protein K"
 FT complement (14799..16418)
 FT /tag= l
 FT /product= spnj
 FT /note= "spinosyn biosynthesis protein J"
 FT 16556..17743
 FT /tag= m
 FT /product= spni
 FT /note= "spinosyn biosynthesis protein I"
 FT complement (17749..18501)
 FT /tag= n
 FT /product= spnh
 FT /note= "spinosyn biosynthesis protein H"
 FT complement (18541..19713)
 FT /tag= o
 FT /product= spng
 FT /note= "spinosyn biosynthesis protein G"
 FT 20168..20995
 FT /tag= p
 FT /product= spnf
 FT /note= "spinosyn biosynthesis protein F"
 FT 21111..28898
 FT /tag= q
 FT /product= spna
 FT /note= "spinosyn biosynthesis protein A"
 FT /transl_except= (Pos:26940..26942, aa:Pro)
 FT 21126..22379
 FT /tag= r
 FT /note= "Beta-ketosynthase domain (KS1): part of the
 FT initiator module"
 FT 22692..23669
 FT /tag= s
 FT /note= "Acyl transferase domain (AT1): part of the
 FT initiator module"
 FT 23793..24041
 FT /tag= t
 FT /note= "Acyl carrier protein domain (ACPI): part of the
 FT msc_feature

DR WPI; 2000-638204/61.
 XX Enhancing resistance of plants against pathogens e.g. fungi, bacteria,
 PT viruses and nematodes, involves incorporating a DNA construct encoding
 PT polyketide synthase into the plant genome to increase defense-related
 PT protein levels
 XX Claim 6; Page 50-52; 68pp; English.
 XX
 CC Disease in plants is caused by biotic and abiotic causes. Biotic
 CC causes include fungi, viruses, bacteria and nematodes and of these
 CC fungi are the most causative agent of disease on plants. A host of
 CC cellular processes enable plants to defend themselves from disease
 CC caused by pathogenic agents. In addition to the localised
 CC hypersensitive response, plants have evolved a systemic defense
 CC mechanism that reduces the impact of subsequent pathogen attacks.
 CC The accumulation of a group of extracellular proteins called
 CC pathogenesis related (PR) proteins were reported to correlate with
 CC the onset of the systemic acquired resistance. Salicylic acid, a
 CC plant produced phenolic compound was implicated as a signal in
 CC systemic acquired resistance. Thus, new methods of enhancing disease
 CC resistance in a plant involves stably incorporating a DNA construct,
 CC comprising a sequence encoding a polyketide synthase (PS) such as
 CC 6-methylsalicylic acid, to a promoter capable of driving gene
 CC expression, into the plant genome, so that the level of
 CC defense-related protein is increased. Alternatively, stably
 CC transforming a plant cell with a polyketide synthase gene and an
 CC additional gene which encodes a protein that, for example, would
 CC increase the level of substrate for the polyketide synthase or
 CC convert the polyketide synthase from an inactive to an active form.
 CC The method is useful for enhancing disease resistance in plants and
 CC commercially viable crops such as maize, wheat, barley, sorghum,
 CC oats, rye, soybean, Brassica sp., sunflower, safflower, alfalfa,
 CC potato, peanut and cotton. The method enhances resistance against
 CC pathogens including fungi, bacteria, viruses and nematodes, and has
 CC applications in agriculture. Advantageously, polyketide synthase
 CC genes can be used to transform any desired plant.
 XX
 SQ Sequence 7131 BP; 1610 A; 2024 C; 1865 G; 1632 T; 0 other;
 Query Match 43.6%; Score 44; DB 21; Length 7131;
 Best Local Similarity 67.4%; Pred. No. 0.0082;
 Matches 62; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
 QY 10 CGCTCCGCGGCTGCTGATGCGCGGTATCCTGACGACGCTGATGACGAA 69
 DB 5415 CGGTCCAGGGTGTGTTTCAACGCTGCTGCTGACGACGCTAGTATGACGACCA 5474
 QY 70 CCCCCGCGGCTTCCGCGCGGCTATGCGGCC 101
 DB 5475 CTCGCGACGCTTCAACGCGGCTTCTCGACCC 5506
 RESULT 13
 AAD17186
 ID AAD17186 standard; DNA; 125401 BP.
 XX
 AC AAD17186;
 XX
 DT 29-NOV-2001 (first entry)
 XX
 DE Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 KM Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KM antifungal; antibiotic; ds.
 XX
 OS Streptomyces noursei.
 XX
 FH Key Location/Qualifiers
 FT CDS 6337..34771
 FT /*tag= "NysI complete protein"
 FT /product= "NysI complete protein"
 FT CDS 34792..51099

FT /*tag= b
 FT /product= "NysJ protein"
 FT CDS 51155..57355
 FT /*tag= c
 FT /product= "NysK protein"
 FT CDS 57503..58687
 FT /*tag= d
 FT /product= "NysL protein"
 FT CDS complement (58786..58980)
 FT /*tag= e
 FT /product= "NysM protein"
 FT CDS complement (59045..60241)
 FT /*tag= f
 FT /product= "NysN protein"
 FT CDS complement (60238..61296)
 FT /*tag= g
 FT /product= "NysD2 complete protein"
 FT CDS 120628..121308
 FT /*tag= h
 FT /product= "NysR4 (long) protein"
 XX
 PN W0200159126-A2.
 XX
 PD 16-AUG-2001.
 XX
 XX 08-FEB-2001; 2001MO-GB00509.
 XX
 PF 08-FEB-2000; 2000GB-0002840.
 XX
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.
 XX
 XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNPE) SINTEF STRIETELSEN IND TEK FORSK.
 PA (ALPR-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIEGLEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVRIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 DR WPI; 2001-557614/62.
 DR P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX
 PS Claim 1; Page 188-254; 266pp; English.
 XX
 XX The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
 Query Match 43.6%; Score 44; DB 22; Length 125401;
 Best Local Similarity 70.2%; Pred. No. 0.0091;
 Matches 59; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 8 GCGGCTCCGCGGCTGCTGATGCGCGGTATCCTGACGACGCGGCTGATGACGA 67
 DB 33721 GCGGCTGACCGGCTGTGTGTCACACCGCGGCTCTCTGACGACGCGGTACTGACCGGCT 33780
 QY 68 AACCCCGCGCGGCTTCCGCGCGGT 91

FEATURES Kosan Biosciences, Inc. (US)
Location/Qualifiers
1..552
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence of the KR domain of extender module 6 of the epothilone PKS"
1..>552
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD33696.1"
/db_xref="GI:21387991"
/translation="DGYLVYTGGLGLISVAGWLAEGAGHLVTVGRSGAVSARQOT
AAALFAGARVTVARADVADRAQIERILREVYASGMPLRGVYVHAGIIDDGLMQOT
PARFAYVAPKVRGALHLHALREAPLPSFVLVYASGAGLLSGPGGQNVAAATFIDAL
AHHRAQGLPALISDMLFADVGL"

BASE COUNT 70 a 176 c 214 g 92 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 552;
Best Local Similarity 100.0%; Pred. No. 7.8e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGCTTCATATCGCGCGTATCTCGACGACGGCGCTGTA 60
|||||
Db 236 CGGGATGCCGCTCCGGCGCTTCATATCGCGCGTATCTCGACGACGGCGCTGTA 295
|||||

QY 61 TGCAGCAACCCCGCGCGCTTCGCGCGTCTATGCGGCC 101
|||||

Db 296 TGCAGCAACCCCGCGCGCTTCGCGCGTCTATGCGGCC 336
|||||

RESULT 2
AX403011 552 bp DNA linear PAT 07-JUN-2002
LOCUS Sequence 34 from Patent WO0183800.
ACCESSION AX403011
VERSION AX403011.1 GI:21387992
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
1
AUTHORS
Arslanian, R.L., Ashley, G., Frykman, S., Julien, B., Katz, L.,
Khosla, C., Lau, J., Licardi, P.J., Regentini, R., Santl, D. and Tang, L.
TITLE
Heterologous production of polyketides
JOURNAL
Patent: WO 0183800-A 34 08-NOV-2001;
Kosan Biosciences, Inc. (US)
Location/Qualifiers
1..552
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="DNA sequence of the mutated KR domain of extender
module 6 of the novel 9-keto-epothilone PKS"
1..>552
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD33697.1"
/db_xref="GI:21387993"
/translation="DGYLVYTGGLGLISVAGWLAEGAGHLVTVGRSGAVSARQOT
AAALFAGARVTVARADVADRAQIERILREVYASGMPLRGVYVHAGIIDDGLMQOT
PARFAYVAPKVRGALHLHALREAPLPSFVLVYASGAGLLSGPGGQNVAAATFIDAL
AHHRAQGLPALISDMLFADVGL"

BASE COUNT 70 a 177 c 212 g 93 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 552;
Best Local Similarity 100.0%; Pred. No. 7.8e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATGCCGCTCCGGCGCTTCATATCGCGCGTATCTCGACGACGGCGCTGTA 60
|||||

Db 236 CGGGATGCCGCTCCGGCGCTTCATATCGCGCGTATCTCGACGACGGCGCTGTA 295
|||||

QY 61 TGCAGCAACCCCGCGCGCTTCGCGCGTCTATGCGGCC 101
|||||

Db 296 TGCAGCAACCCCGCGCGCTTCGCGCGTCTATGCGGCC 336
|||||

RESULT 3
AF217189 58733 bp DNA linear BCT 09-JUN-2000
LOCUS Sorangium cellulosum putative transposase gene, partial cds;
DEFINITION putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)
Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE
Cloning and heterologous expression of the epothilone gene cluster
JOURNAL
Science 287 (5453), 640-642 (2000)
MEDLINE
20115955
PUBMED
10649995
REFERENCE
2 (bases 1 to 58733)
Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE
Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL
Gene 249 (1-2), 153-160 (2000)
MEDLINE
20293058
PUBMED
10831849
REFERENCE
3 (bases 1 to 58733)
Julien, B.
TITLE
Direct Submission
JOURNAL
Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
Location/Qualifiers
1..58733
/organism="Polyangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1..992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHEAADAGCGPRVLLYNLKNAYVERGDAIRRPITLAL
SDYRFPFPAVAVARGNEGRVERAIRVYREGFEFARAYADIGDINRQATEWTSAL
DSWVEDRARIYVRAAFDDERSVLLRHPDPPEPDHREVEVGKPTPYARDLDYSPH
DTRRTLVVLAIDSOVRIADGNQIVATYHVSRSRQQLQEPDHLRLVDEKRAEHR
GLDRLARARSSQAFRLRIYVARGDNVGSALRLDLQLDVGAALAEALVLEBDTI
HTGAVROVDRDRRSERHLPPPSIVTGEHALVVTTHSLTYALKKDPF"
989..1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTETKDRLSLGLGILLACBEOADKPMULREVIAIERE
RHRSIERLKNRYAAAFKPMTPDPSWPKIDREAVVDLQYSRVDLLEFVYTRRD
AOKPLLSSTNKAFAWMGVFPAACVYTLVDRLVHRAVEIIEAESYRLKEAKELNAT
RTRKRTKKH"
1998..6263
/gene="epoA"
1998..6263

Oy 61 TGCAGCAACCCCGCGGCTTCGCGGCTCATGGCGCC 101
Db 37000 TGCAGCAACCCCGCGGCTTCGCGGCTCATGGCGCC 37040

RESULT 4
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epoethliones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1. 71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2,5e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CGGGATGCCGCTCGCGGCTGTCATGCGCGGCTATCTGACGACGGCTGCTGA 60
Db 36940 CGGGATGCCGCTCGCGGCTGTCATGCGCGGCTATCTGACGACGGCTGCTGA 36999

Oy 61 TGCAGCAACCCCGCGGCTTCGCGGCTCATGGCGCC 101
Db 37000 TGCAGCAACCCCGCGGCTTCGCGGCTCATGGCGCC 37040

RESULT 5
AF210843 68750 bp DNA linear BCT 21-JAN-2000
LOCUS AF210843
DEFINITION Sorangium cellulosum strain So ce90 epoethlione biosynthesis gene
ACCESSION AF210843
KEYWORDS AF210843.1 GI:6724237
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
Molnar,I., Schnupp,T., Ono,M., Zirkle,R., Milnamow,M., Cyr,D.D.,
Goriach,J., Mayo,J.M., Hu,A., Goff,S., Schmidt,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing
agents epoethliones A and B from Sorangium cellulosum So ce90
JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultures Research Institute, Inc., 3054 Cornwallis Rd., P.O.Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source 1. 68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
CDS <1..1826
/note="ORF1"
/code_start=3

/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASDALFARATSAVYLDDGHGATPRHYLAERAGIEDLRALREH
LRIQEGPSFHCMLGDLTVELLADOPILASTSFHARRLRHPDMTSDAMLVGDAVL
RNILNAGAPLREYEERERARTAOEARRLWLAAPCFADLPFEDDAGLPLGP
MSPEVAERRLRASYATPELACALMLWLGAGPWSGYPAYEMLPENLLFGILPT
ALAAASPGTSEALRGARLPASMEVSSKKSQGNIEALMERLRTIVRAMGNAD
LSRPERAEALAAEVRRLRAOPAPPAAGALAVAGVSSGRLSGLYTDGALYSGDND
LYMPGRISPVYLACTGDPFEELAPISOMLFVAHANAGTISKYLTGSPILVARN
QAPMSLYARCPMAVNQAMPDPERCAFPYVQNSTIIEFHPPLPRCLHEPAGSARF
LACDEHLTWCELSAGRIELMRHPHRRPGAPSRFAYLGEHPILAAATWPSYLTANATHV
MADPBRALIVGDKRTGVEPIVLAETRHPPHVSERDRIFFALITQDPSRDHVEHIR
SCASTVADYOROLMDRPMVLRRLGFTFTMDRLITLARS"
complement(1900..3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to Dd-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAPGALGAEESQPGODAGALAPAREYMAEVA
ACOMPGAVMLVARGDVHDVAVTELGSGAPMRDITFRISMTRKATVATVAMLVE
ECKLDDSPVDRLPELARKVLRIDGIDETVPAERITVRDLMFTMGISFDA
SSPIORALDELGLVNAQVPVMTPHGDWEIRRLGTLPLMHOGAQMVTGSLGV
VRRADQGRDFAVRERILAPLGMRODPFHVPAADKARFAGCCGCTDEQGTREARD
GASATSPAPPSGAGLVSTVDITLFAKMLMGVHEGRLLSASVSRMTADHL
TPAQKAASSFFPFGFETHGWGVMAVVAADVAEVSEPGRYGMDGFGISWIDPDEL
IGIVMOSAGFLFSGALERFMRSVYVAESA"
3415..5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGTFEROVLSTVLALIVTARSGELARRLOPEVIGELFG
GVLPSPVYGALAPGPHRALROEPANGVLSGISWGLALLLMGITEVDGILKEKA
RGAISALCAIAPPLAAGAFSALVLDRLPBGPLFGILYSTANSVAKYLIERESM
RRTSVQNTLAAGVSEVAWVLAAMTSSYSGSPALAAVRSALASGFLVNLVGR
LPHLAMRWVADATRVSKGQVSLVLTFLAALTORGLHPLGLAFALVNLNSPRT
NRPILDGVQTLVAFAPVFEVLGAKRVVDSOLRPAAGVTALLATLTAKVAPAA
LGARLGGRLGSEALVAVGLNMGGDLVVALVYGLGSEALVATMAVVALVYTA
SPALLIWEKRRAPQOESARLREBEARRAYIPGERTIVPVAHALDGFATDYES
IVASRRKGETVDITELSYEQAPGFSRAGASRLARLAVGIRKQRELRGS
IOAILRASRDHLVIGARSPARAKMSFGRLQDNLVORAESNVLVVGDPPAAERAS
ARRILVPIIGLEYSFAADLAHVALWMAEVLVLSAQTDGAVVWRDRERSRVAV
ARSVDVAVFRGRRLGVRVSRVHGAAHPSDEITRELAARAPYDLVLGCGYDHGPIGRL
YGSYTESVVSVRVVALLVHAGGREGVR"
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGTRCAIABSYITGGSNGPEPQMTSHEYACLLNASRDA
OVAIVYFSDRDPAGPYRVTPARRTRHVRNDLIEPEIPDDTIVASIESDAPLIV
QHTRLDSRQENALLSTIAYIDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHYDTGRFRGRIGHTLGLASMLACGGSPSEKTVQTRLAP
GADARYTADVPDDATTRLAIVYVHLSPPERLEASGERVYVQRPSPESPWRVRYGLD
IVADSRGLAETVYVAFNEFLITRAKSSQSPSPSAVIVTPTSGV"
7610..11875

CDS

/gene="epoa"
7610..11875
/gene="epoa"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADPTEERAEDPDIATVGASCRPGVIDSGFWILLEGSDRY
GRVPAERMDAAMPDDPDPAKPTPTTRASTLSVACFDASFEGISPREALRMPAHR
LLEVEWEALEMAIAPSALVGETGVFVIGPEYPAALPQATSAE IDAHGIGTM
PSVAGRIATYALGIRPCVAVDTETSSSLVAHACOSLSRGECSTALAGVSLMLP
STVWLKSTKTRALARDGRCAKFADEGREGGAAVVLKRLSGARDDGGLIIVNG
SAIHNDGASGGITVNGSSOEIVLKRALADGCAASGVYEAGHTGTGDDPTEIOA
LNAVYGGRPVATPPLIGSKYTNLGHPEYASGITGLKLVLSLOHGOI PAHLAOLN
PRIWGDRLVTRKATPMDMTPRRAGVSSFGMSGTNAHVLEEAATCTPPAE
RPAPLLVLSARTASALDOAARLMDLETYPSCLDGVASLARTSMEHRILVAAT
SRERLRAALDAAAGQTPGAVRSIADSSRGKLAFLFTGGGAOTLGMKRGIVDWSAF
REAFDLCVRLFNOLDRPLREVMAEPAVDALDQTAFTOPALFTFEYVALALMS
MGVEPELVAGHSIGELVACVAGVFSLDPAVETVVAARGRLMOALPAGAGVSTRAPA
DYAAAVAFHASVSIANAVPDOVYTAGAOPVHAITAAAMARCAKRLHVSATFS
PLAAMPLEAFGRVAESVYRPSIVLVSNLGGKCTDEVSPGIVWHAREVFAAG
VKALHAGAGTFEVEGKSTLLGLVPACMDPARPALASAGDEBATAVLEALGDM
AVGGLVSMAGLTFPSGGRRPVLPYPMQERERWIDTKADDAAGDRAPAGAHDEEAG
GAVRGDRSARLDHPPESGRREKVEAAGRPRLDEIDEGVLDHILVRYTERARAG
LGVEIVDAAGLSFNDVOLAGVPPDLGGKPMPLILGEGCAGRIVANGVNGVY
VGQPVILASAGARTHTTSAALVLRPQALSAIEAAMPVATVLTATYALDRILRDP
GERVLIHAATGGVGLAIVQAQHVGAEVHATAGVEKRALESIGVYVSDSRDRV
ADVMAATGGGVDVNLNSGELIDKSFNLRSGRVELEKRCYADNDQGLRPLR
NLSESLVDKMMLEPARVRLLEELIGLAAGVFPPIATPIARVADAFSMAG
AOHLGKLVLTGDEVOIRIPTHAGAGPTGDRDLRLRLASAAAPAAALAEALRQ
VSOVLRPEIKVGALEFTRLGMSIDMAVELRNIEATKIKLSTTSLSPNALLA
ONILDALATALSERVAENLRAGVQNDPVSAGADQMEITAL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P, contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQILNELHOGIKLAADGERLQIOAPKNALPNILARISEH
KSTILTLRORLPAESIVPAERHAPFPLDIOESSTGTFVYPSGILHAREDD
CTDDVRLSRAPKVVYARHMDLRANTLIDITRKSSEORAKRIQEDLWAMHBCDVS
RIYSLDAMSHRIYDTPREPLHYVAVRLDEROTRLISDLINVDGSLSTFKMDL
SYEDPETSIPVLELSTRDYVLALESKKSEARORSMDYKRIIAELPPPTLPKAD
PSTLKEIRFHTEDWLPDSMGRUKRVEGERGLPTGVIIAASEVIGRAMSASRFTL
NITLFRNLPVHPRYNDITGDTSMVLDITTRKSEORAKRIQEDLWAMHBCDVS
GIEVOREAAVLGIORGALPPLYLTSLGVVTSLQRLGTFVYSTQTPOLDLH
OLYEHDDGLVLAWDIVDGFPPDLIDMLAAVYVFLRLLEEPGGEQVORPAOLE
ARASANTNALISEHTLHGLFAARVLOLAVVSAKRTIYEELSRRSRRLCARL
BOGARPTIIVAVMEKGEVYAVLANLEGAAYVPIADLPAERHTYLLDHGCEKLV
LTQPWLDGKLSMPGIDRLVSEAGVGGDDQPPMHIQTPSDIAYVITVSGTGLPK
GVMIDHGAVALNITLIDINEREIIPGQDVLAISLSFSLSYIVDFGIIAAGGTVVPA
SKLRDPAHMAELIEREKVYWNISVPAALRMVLVEHFEGRPDSIARSLRLSLSGMWIVY

GLPGLQAIRPGVSVISLGATEASINSIGYPVNVNLSMASIPIYGRPLRQTFHVL
EALPRPVWVGOLYIGGVGALGYCMMDEEKTARKSLFVHPTEGRKLTGDLGYLPD
GNIFPMGEDNOIKRLGRVLEIGTIEETKSHRPVAVRPIVGVNDANKLLAYVVP
EGTRRRAEODASIKTERIARAAHAAADGSLDEROFTKARGLRDLDRKPVNL
TGDPREGLDVITARRSVRTLEAPLPFVEGFRSLCSLSEVEDGATLPEKFRPSAG
STYPVOYAAVKSRIEGVDEGEFYHPEFERLLKLSDHGIERAHARQNDVFEDEA
FNLLFVRIDIAIESLSRSSRRCFLIAGYMAOLLMCAQASCNGVPOQFNEQVRA
PVLDRSHDDVYHGMIGRVDPRQFOCTLGODSSPRATRGAPPCREQHFADMLRD
LFRKLPEVWPYTFVLELDLPLTSNCKVDRKALREKRDTPSPHSGHTAPRALREI
LVAVRYELGLEVYGLQOSFYDLCAITSIHYRMSLLQKRLREIATLTELFOYNLGS
LASVLRDSRDLDGRPNMDREVERRRGRRS"
16251..21749
/gene="epob"
/gene="epob"
16251..21749
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match 98.4%; Score 99.4; DB 1; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

gene

CDS

1 CGGGATGCCGCTCCGGCGGTCGTTATGCGCGGATCTGGACAGCGGCTGCTGA 60
|||||
42552 CGGGATGCCGCTCCGGCGGTCGTTATGCGCGGATCTGGACAGCGGCTGCTGA 42611
|||||

61 TGCAGCAAAACCCCGCGGCTTCGCGGCGATCGCCGCC 101
|||||
42612 TGCAGCAAAACCCCGCGGCTTCGCGGCGATCGCCGCC 42652
|||||

RESULT 6
AR193029 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR193029
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ilyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CGGGATGCCGCTCCGGCGGTCGTTATGCGCGGATCTGGACAGCGGCTGCTGA 60
|||||
42552 CGGGATGCCGCTCCGGCGGTCGTTATGCGCGGATCTGGACAGCGGCTGCTGA 42611
|||||

61 TGCAGCAAAACCCCGCGGCTTCGCGGCGATCGCCGCC 101
|||||
42612 TGCAGCAAAACCCCGCGGCTTCGCGGCGATCGCCGCC 42652
|||||

RESULT 7
AR199551 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199551
DEFINITION Sequence 1 from patent US 6355457.

ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5,4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGAATCCTGAGACGAGGGCTGCTGA 42611
|||||

Qy 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652
|||||

RESULT 8
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5,4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGAATCCTGAGACGAGGGCTGCTGA 42611
|||||

Qy 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652
|||||

RESULT 9
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5,4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGAATCCTGAGACGAGGGCTGCTGA 42611
|||||

Qy 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652
|||||

RESULT 10
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5,4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGATCCTGAGACGAGGGCTGCTGA 60
|||||
Db 42552 CGGGATGCCGCTCCGGCGCTGCTCATGCGCGCGAATCCTGAGACGAGGGCTGCTGA 42611
|||||

Qy 61 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGCTTCGCGCGGTCATGCGGCC 42652
|||||

RESULT 11
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones

JOURNAL Patent: US 638737-A 1 07-MAY-2002;
FEATURES Location/Qualifiers
SOURCE 1. 68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25359 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.4e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGGATCCCGCTCCGCGCGCTGCTTCATGCGCGCGGTATCTCGAGCAGCGGGCTGCTGA 60
|||||
Db 42552 CGGGGATCCCGCTCCGCGCGCTGCTTCATGCGCGCGGTATCTCGAGCAGCGGGCTGCTGA 42611
|||||

QY 61 TGCAGCAAAACCCCGCGGGTTCGCGCGCGGTATGCGGCC 101
|||||
Db 42612 TGCAGCAAAACCCCGCGGGTTCGCGCGCGGTATGCGGCC 42652
|||||

RESULT 12
AX024384 16124 bp DNA linear BCT 15-SEP-2000
LOCUS
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer, S. and Mueller, R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOLOGICAL FORSCHUNG GMBH (DE)
FEATURES Location/Qualifiers
1. 16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN

Query Match 75.2%; Score 76; DB 1; Length 16124;
Best Local Similarity 85.0%; Pred. No. 6.4e-07;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CGGGGATCCCGCTCCGCGCGCTGCTTCATGCGCGCGGTATCTCGAGCAGCGGGCTGCTGA 60
|||||
Db 5317 CGGGGATCCCGCTCCGCGCGCTGCTTCATGCGCGCGGTATCTCGAGCAGCGGGCTGCTGA 5376
|||||

QY 61 TGCAGCAAAACCCCGCGGGTTCGCGCGCGGTATGCGGCC 100
|||||
Db 5377 TGCAGCAAAACCCCGCGGGTTCGCGCGCGGTATGCGGCC 5416
|||||

RESULT 13
AX024277 16124 bp DNA linear PAT 15-SEP-2000
LOCUS
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer, S. and Mueller, R.J.
JOURNAL Patent: DE 19846493-A 82 13-APR-2000;
BIOLOGICAL FORSCHUNG GMBH (DE)
FEATURES Location/Qualifiers
1. 16124
/organism="Polyangium cellulosum"

/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN

Query Match 75.2%; Score 76; DB 6; Length 16124;
Best Local Similarity 85.0%; Pred. No. 6.4e-07;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 CGGGGATCCCGCTCCGCGCGCTGCTTCATGCGCGCGGTATCTCGAGCAGCGGGCTGCTGA 60
|||||
Db 5317 CGGGGATCCCGCTCCGCGCGCTGCTTCATGCGCGCGGTATCTCGAGCAGCGGGCTGCTGA 5376
|||||

QY 61 TGCAGCAAAACCCCGCGGGTTCGCGCGCGGTATGCGGCC 100
|||||
Db 5377 TGCAGCAAAACCCCGCGGGTTCGCGCGCGGTATGCGGCC 5416
|||||

RESULT 14
AF319998/c 49736 bp DNA linear BCT 28-MAY-2001
LOCUS
DEFINITION Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
complete sequence.
ACCESSION AF319998
VERSION AF319998.1 GI:14210834
KEYWORDS
SOURCE Stigmatella aurantiaca.
ORGANISM Stigmatella aurantiaca
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Cytophactinaceae; Cytophactinaceae; Stigmatella.
REFERENCE 1 (bases 1 to 49736)
AUTHORS Sliakowski, B., Nordstieck, G., Kunze, B., Bloeker, H. and Muller, R.
JOURNAL Novel features in a combined polyketide synthase/non-ribosomal
peptide synthetase: the myxalamid biosynthetic gene cluster of the
myxobacterium Stigmatella aurantiaca Sgals
Chem. Biol. 8 (1), 59-69 (2001)
MEDLINE 21110452
PUBMED 11182319

REFERENCE 2 (bases 1 to 49736)
AUTHORS Sliakowski, B., Nordstieck, G., Bloeker, H. and Mueller, R.
JOURNAL Direct Submission
TITLE Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
JOURNAL 38124, Germany
FEATURES Location/Qualifiers
1. 49736
/organism="Stigmatella aurantiaca"
/db_xref="taxon:41"
/complement(2. 1306)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="putative aspartokinase"
/protein_id="AAK57182.1"
/db_xref="GI:14210835"
/translation="MRPISARQAGRGTSALALIVOKYSGTSVDERIKNARKCIA
AOKAGHDVVVVVSSMGETNRLIKVIOITERPERODVYVAVAGVSGVLAALIO
TOKRAVSEFGHOVIRITDSTFSKARIKSIDARINYALKKNNIVVAVAGVSGO
VTILGRGSGDTTAYALAAKADCEITTYVDGYTTDPMMVPAARLEIAIEBME
LASVAGARVLIIRSEVEMARKYKVPDLMWVSFTDDGTVLCEDAMEDEVLSGLATDK
EAKIAIRGVDPVGVAAKIRGALDEOSIYVDLIVONYSKGRDVFETVVKADLTAK
DVYKIKASVKAEBEVDQVSKVIVGVAMRNHSGVAAAFVLAGAGVNVOMISTS
EIKVSCVHSNTELAVALRSHTAFGLDKPAGAVSEVSLKG"
1363. 2895
/note="ORF2"
/codon_start=1
/transl_table=11
/product="putative histidine ammonium lyase"
/protein_id="AAK57183.1"
/db_xref="GI:14210836"
/translation="MSRRPLNTDSDITLKEETIIQVARRHTVTELAPAAARVAAARD
VDRVAADPTSYGINTGETGLAEVRIDKRLREIQRNLILSHAGVSGPLDEARVL
LLRLCNVLAAGYSIGINTGETGLAEVRIDKRLREIQRNLILSHAGVSGPLDEARVL
GEEFAFKGERLPAALALERAGIKPVVLEAKREGIALVNGQAMCAVGTLLQALAEMLA
DLADLAGAMTLEGLISHKRPFEIDQVRAHEQKACAAHLREILDALSALVESHVNS

```

gene
CDS
    complement(3275..7822)
    /gene="mxa1"
    complement(3275..7822)
    /note="non-ribosomal peptide synthetase; contains a
    condensing, adenylation, and PCP reductase domains"
    /codon_start=1
    /transl_table=11
    /product="MxaA"
    /protein_id="AAK57184.1"
    /db_xref="GI:14210837"
    /translation="MSVRELLTTLIAKGYRLRLEGGDKLOVEASKAVLPPELRBAELSAN
    KADIIAFLKXHSREETSGLQPLARRPVITLSSGOERLMLFDRUSPETSOTYNHLG
    LKARGALDKLRSLDELVRHVELRTHPEVEGSPROIVSPDTGVALITVELLGL
    TESQREEEIYRSDLEAQQPDLAKGPLRLVYLAGPDFALEFYTKHHIITDGLSLG
    VVEERLSMLYASFVGGQAALPPVSKQFADSLREBAMLAGSGARERYMKGKLLG
    PPLQLPVDAVSTSHRGATVPALSPALSDALREIAREGCSFEFTVFAAFSALLH
    RYSGVDRCGVIVANKRGVPADILGFIANTALRCDSGEPTESQMLARAKVYLEA
    LDYQALPSEVQIVGASRDGGLNPLVRCFTLESTIPATLDPGTSWFLNGADGS
    VSGAKFELSLILASERGLACMLEYSRVFPASTYERWVGHFOVLESIVAHPEVPL
    SKPLPITAEERGLNDMNGPVLDPALCMHBLVOAERTPOAIVAVVSGQRTLYAE
    LNRANQALAHHLRLGIGIOKEEREGLCVERTEIVIGLAILKAGGAYVLDPAVPER
    LALILEDQVPVLTLOORLVPRLPATQAVVCLDKDMPITGASDNPRIATVAPLEA
    VLIYTSGSGTKRKVMIEHRNVAFLIMAMVSFSPRELAGTIASTISICDSVPELFT
    PLCCGAKYIVAKNALPELPLPARETTLINIVPSAMGALLRSGVPSAIVNLGEEA
    LAGALVDQIYQLDHVDVFNLYGPSERTTYSTFTNRNGQPTTIGRPVNTQVYLA
    NNEPMPIGVGVYIGGMVARGYLGRELTAEERYRSPFGGSEARLRYTGDILKML
    PGQOLEYIGRMHQVLRGFRLEIGELGAVLMEHSGIRDAVVVRELGADQOLVAVY
    VGRGKAPAPALRDYLRKSLPEYVWPLFVGLDALPLTPNGKVDRAALPAPERHSG
    PAKHEVAPTPGEESLAIWROVLGVEOGADNPFELCGHSLILYRVLVARSASGA
    DIPRLALDQPTLEEMARAVEAKTGSLEADHYTMEADAVLDEIALGRLPEPTG
    ALRTLLTGATGFLGAFLEELCRRTDARTYCLVRSKTEQEGEMNRLRKLNEYSIMNE
    ALAPRIYPRGDIQPLGLSEKEFORLSEEDAIYHNGALNPLPYPSMAANVAGL
    TREIRLATRTIRIKPLHYSTVSVPLGRKAIREDPELEGSSLYGGAQSKWAEK
    LVREASRRLPVTILRPGRTVGTSHRGANTDDVLCRTLKGCVRMGVAPSDALLST
    PVDVYSATVDSMRPESTGOTYLVNQFVRADMMNVMRAFGYGLRVLPLYDMLSE
    LGSAASSDSELDLMLFLQQVPPEDRSVGGPRMYVCDSDTTLKALGGTSCPSVDAS
    LSTYLSLVHNGFLKAPVR"
    complement(7819..11520)
    /gene="mxa1"
    complement(7819..11520)
    /note="PKS; ketosynthase; acyltransferase"
    /codon_start=1
    /transl_table=11
    /product="MxaB1"
    /protein_id="AAK57185.1"
    /db_xref="GI:14210838"
    /translation="MRAQGGHPILGVSFTTSYQKENTRWESALSKKPSYIADHRVY
    GAVVYPGAYLEMLSAARELSGEKAYELVDTSFKEGLLPGEQEVVOMALSDEGD
    ALSTQLSQKATSEPNAPAGMAHVAGQIRTPTEGGSTQTPSPFEAIRKCEVYLAQA
    AHYEELASGVAYGPAFOGVQOIMRGRAEALGHLRPEPLARASARYRHPALLACF
    QLVAAVPOEDLTDPAGPAPVALAHLRLEHPEGTEFCHVLRPSEGRKRVYEDL
    VLRDASGRILVEALIGLRVOOLDAPAAEHAGKDLFPALEENRHAASVOOVYABGASGR
    WLLINDGSEIADTYSQLEBERGEYTRNDLAGAGGRVYDNTSQOAPGCVYSBAPGA
    GSAPRGVILHTGLEPREVDPLMGCGVHLVVALSRVSLASPRMLLVTRGVHHSGS
    RSASOVASLWGLGRTLAHERSELKTRVDVSAAGMGDAGALMLRELLVADDEEI
    SLRADGRHYVRIGRSPGRAPSEAVYPAGPGFHLERDQGRLEFRKARPPGGEI
    EIEVEVALGSGEGTEPERMEAGEGIGCGRVVAGIGEVNVAAGEIRLIVAGL
    GSHVSPVAGCTAIRSSSSHGEBALAANTVYIPAWALPHTGLERKRGVATVAGSGL
    GRAAIKLAKRAGAEYFAYATTDQRASLRELCADVMDPSDSEYSRVADTGGKVE
    LAIVAGAVELERSLPTLSEGARVLDLRSGASARTGDANAMCVADVDQFVRRRPER
    FASLQEVKAEEAGGLPASAADAKLKDARAVVYPADARRRLDGTVLVGGEL
    GGLGLAVAKMVEOGAGHLVLYGRDVTYSEQOEVTALAEAGARVARAVASVRAQ
    IARILSEIAEGGAPLRGYIHAAGVLDGGLVLTQOYTERPRVAPAPVLYLGGWNLVJRE
    APIDFEVLYSSAAMEGAPCGGNVYAGNAFLDALHHRALGLPGLSTINMGFSVEYL
    AAQANRGERLQAQSGSDSTITPAGNTILGRLLDGDVQTAWAPLELRKAVETYPKRS
    SPWLSLVPADAGVEGVARVALLTLETTATVPBAKKAALFVREOLAARYLDSAR

```

```

gene
CDS
    complement(11487..14303)
    /gene="mxa2"
    complement(11487..14303)
    /gene="mxa2"
    /note="PKS; dehydratase; enoylreductase; ketoreductase;
    ACP"
    /codon_start=1
    /transl_table=11
    /product="MxaB2"
    /protein_id="AAK57186.1"
    /db_xref="GI:14210839"
    /translation="MSKQEPAPQPSALORALVEKRNQARLDAVERARTEPIAVTGL
    GCRPGAVDSSGYWKILRDGDVALREVPESWDVPGHEDSTRGPGMYGRRGFLD
    DYEHFDLPEFGISPREASLDPQORLTVLEAMEALENGLAPDOLAGSKTQVGMVMS
    SRYMARLKENATATPDGVMATGNCTSYTPRGVSYVLGQCFMVPDTRACSSSYIAH
    LASESLRGEESMLALAGVNLILSPETMCLCSQGLVLPNPKCTFDASADYVYBEG
    CGVILKRLSDQRDGSILALIRGSANHDSAGALVPPNPGSOQAAVORALDMARI
    APALVYIEAHGTGTPLDPIELRALGAVLGRPADRPFPGSYKTMIGHLEPAGI
    AGVITLTILOKEIIPHLHFTPPHWEMDRIIPARVPERVPMFAGHGRITAGVSEF
    GLSGINAHVYLEEAPASTAPTAVEDKAOVLVYSAKTPRALSSIAEAMSSPTKMEG
    GSIADISTSAISRTPHERRIALAKSKQELIQLQEARAKKSHAGLSSGRRGTGQ
    KIVFVPGQSQMLQGLTAEVNTREDAIDWCHESLRTDMSLRDLTPNGQQR
    NMDIDVIOPTTEALQVLAALAAVWRMLGLEPFAVHVGSMGEVAAHAAGALSLEDAARI
    CERSKILRVSQCGAMAVDLSREQAEEIRKQEDRIIAVANSGRSTVLSGDPALAK
    EYLERLERREYRCRWKVDVASHSQMDPLRODILDRMAALIPGSAVPLISTVYAG
    IDGRGLSAPYWRNLROPVLPFADAVQRLIDQGNVYFTELSPPRLVLPVDMALRGDT
    SGRGLVSVSLITDQCEMPSILSSVQALTTRYIAWVGSGYIPISVAGLRFRLTRGASG
    TQMSQOQLRRRSCALAREVIPS"
    complement(14300..29878)
    /gene="mxa2"
    complement(14300..29878)
    /note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
    ACP regions"
    /codon_start=1
    /transl_table=11
    /product="MxaC"
    /protein_id="AAK57187.1"
    /db_xref="GI:14210840"
    /translation="MANRNDESNEISPAKRMILVALEKQTRLNAVEGAKEPIATVGM
    ACRPOADVNEAFWRLLQGVDAVREIPLERKQDVNDLIGGAKRWGFLDSVDSFD
    PGFEGISPREAVRMDPOORLLLEVAMELEDDQKDEVDLSGSRVCFIGAENDVDHGM
    QYVERPETGAFSATGVAASVLSGSLSYFNLOGPLRVKAVDTACSSLTLGACISLRN
    RECNALAGVNLILSPOSVLVSLKQALSPGRKADASANGSTVREGGCIYVLRK
    LSDALADGHIATIRGSAIINDQGSTGLTPPNVLSOALIRLOALESSGIAEOVSYV
    EAHGTGTPLDGPIEALRETYGVPRPGSGVCGISYVTTNNGHLESNAGIAGIKVYL
    AMRHOSITPPLHLKQVNPRTQLEGSALTTPIRLTWKASGQRRRAGVSSFJSGTNAH
    VITEEAPAVAKDALPQREPVLPJSAREREAVALAERYAEKLRGEGGSLSDVCYTA
    SAHRTHTHRLVALDSREHTEERLARARAGQPGSGARAKFGRKRIYFVYPGQG
    AORAGMGRLMESEVPFDALVYKDKALQVDEHVAAGLDEIDAEORSLKAEFSDQCKGMV
    LEFAVEVLTALMRSGVIEPDVAMGSMGEVAAHAAGALSLEDALEITICHSKLMTL
    GGRCATMANVELSYEBEATIRPNWKGITHLGSLNRPROYLSGEPAYKEAVEALGRK
    GVFTRWVNRVPRSSHPLVEVLSIRELGRITVTPRAADIPICSTVTGFLDGAKIDAA
    HNGNLLRVNAPVAAAOHLASDRETEFVESHAPVLAPEVDETMAGTVEGYALSLR
    RNEDRHRMTLSLGLYSOGVANNRNYVPGQGLVPLTPYPMQERHWMPEAAVAGV
    ASRTSGHPLIERIDLSNOGVSRLVQOLGKPLPLAEHRVGVADLPAALAAADM
    ALSAKQITLGEVSVLEDSYSEFSLSEFPDEEVRLOJMLSTERPGASARFLAPAPRE
    TPSDELNTALHAEGSLRNGDASVYKAPVALDELRACGEBASVEAHYQTLKEIGLE
    TGPSTRVARELIRANGELDGRVQGEGLTAEERI SLHPALDQCFQVLAALBGLDS
    BEGVVYPALSLRLVLRPEALMHSHVITVSAEGMSVTIEDVTLINDQORVAAEV
    RLVCKRIIAASAPADEVVRMITYSAQEQOORHGRVETSSSTRAGMVLPEADARGIGOS
    LASGILAGGCGVVEEADGVSRYRLRADRYRINPSKLEDFOLALKEAFSDQCKGMV
    LYGLDVAANVMEPESEFEALSTISSTVYLIQANRQRMQRPRLMTVTVQVYIAGS
    EOPGSPAPALVGLAAVITTHEPELRCHAHVDSADQSPDEETASLTELIGCAEDRIA
    LKGAVRHARLGRIGVLSGRKAAEPREFTVADGYLLITGGIGGLGLEVAKMF IORGANLNA
    LTRGAPAREEADKRGVIGELRANGAEVPRVADVTASDVGARILGSLIDSLPLRGVNA
    AGVINDGYLTQDEPRLEAVLAPKQVGMNHLVQOGLQPLETFVMESSAASMGSPGQ
    GNYAANAMPQVLDLHHRRAAGIPGLSGAMNHPADVGLAAVNTNRDRLALALGMPG
    AQAIDALIRLVAQEAQVAVYDILRQMRREPLSAQSPSLSYLMQEVSVSPASKGS
    ARELILAAEPQGRKIGLSYLBQGLRLKREPSLEDPQPGDGLGMSLGLLELRNK
    IESGIKTLISATLIYAVPYOALTLYYKGLPEVYLEAERAPTVRPTQREPRVSGVAP

```

Query Match 52.5%; Score 53; DB 1; Length 49736;
Best Local Similarity .73.1%; Pred. No. 0.033;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 9 CCGCTCCGCGCGCTGTCATGCGCGCCGATATCTGAGACGCGCTGTCATGACGCA 68
Db 8739 CCGCTGAGAGCGGTCATGCGCGCGCGGCTGTGTGAGACGAGGCTCTGTCAGACG 8680

Qy 69 ACCCGCGCGGCTCCGCGCGCTGTCATGCGCGCC 101
Db 8679 ACGGTGAAGCGGTCGCGCGCTGATGCGCC 8647

RESULT 15
AF232752 3225 bp DNA linear BCT 15-FEB-2002
DEFINITION Mycobacterium avium subsp. paratuberculosis polyketide synthase 7
ACCESSION AF232752
VERSION AF232752.1 GI:8927424
KEYWORDS
SOURCE Mycobacterium avium subsp. paratuberculosis.
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
1 (bases 1 to 3225)
Bannantine,J.P. and Stabel,J.R.
Identification of two Mycobacterium avium subspecies
paratuberculosis gene products differentially recognised by sera
from rabbits immunised with live mycobacteria but not heat-killed
mycobacteria
J. Med. Microbiol. 50 (9), 795-804 (2001)

JOURNAL MEDLINE 21432816
PUBMED 11549181
REFERENCE 2 (bases 1 to 3225)
AUTHORS Bannantine,J.P.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Bacterial Diseases of Livestock, National
Animal Disease Center, 2300 N. Dayton Ave., Ames, IA 50010, USA
FEATURES
source 1..3225
/organism="Mycobacterium avium subsp. paratuberculosis"
/sub_species="paratuberculosis"
/db_xref="taxon:1770"
<1..2038
/note="Pk57"
/codon_start=2
/transl_table=11
/product="polyketide synthase 7"
/protein_id="AAF82076.1"
/db_xref="GI:8927425"
/translation="DORLTAVPPGLSLVAAGVYVELTALYGLSVLAGLPGSERVL
VHTATGGVMAVOLARHYGAEVAFATASRGKMDLIRAMGEDDDHISDSPDEFKFL
AATGGAGVDVILNLSAGEFTDASRLIAPGRFTEMKTDVDPDVAQRGARARYA
FDLMEAGPDRTRAMLAETVGLDQGVILPLPKTFDVCASATRYVSQAHIGRVYL
TVPSGLDEVLSGCGGSLAQSTVLLTGGGAGSALARLVDYRVGHVILSRGAQA
AGAEVLDELORAGASAYLACDVADRDVATMLAGLPAGYPLRGIVHAAGILDDGL
SSLIPRDVAVIRAKVDGAMNHELTDLSAFVFSMAGIVTGPQANYAANSF
LDGLVARRADGLAGLSIANGIMEOASMTAHLGDRDKARMSRTGLAPISPEOALAF
DAAMLVETPVIVARLDRALSENIALPLIRELAGPFRVITDDADYTRASGGLAA
RLHGLSPARRRELVDLVCGNAAMVILGVPNADINAGAFODLGFDSLTAVELNRUK
NATGLTSPILFDYPTIPVIVLAHLDSRLAGSGDDQDPLMGRENDDITRELQALLGAA
HMNSDDKAVLRTIRHGLGALPADGPDSPALDDELAEATESQLPAILDELGR"
2058..>3225
/note="Pk58"
/codon_start=1
/transl_table=11
/product="polyketide synthase 8"
/protein_id="AAF82077.1"
/db_xref="GI:8927426"
/translation="MPGTGHLIDYIKRLTADLRTRRRVALEBGRLESPAVVGMACR

YAGVDSPEALMDLVIEGRDTSDFPYDRGMVDEGLYDPPDAKGMKTYRQGSFLQHA
GDFPAGFEIGISPEALAMDPOORTMLEICWELRABAGIDPSALRGATGYAGYTHAG
YGEVKEIEBEGYGLTSGTILSYTSGRVSIVGLIEPAYSVDPAACSSIVAMHIAOSIR
SGECDLALAGVTVMAIPATVRSRQALAPDRCKYRAGAAIGTSBSAGVLYLE
RLGDARRIGHPVMAVILRGSAAVNDGASNGLTPNPGSQORVIRALANAGLSAVDVY
VEGHTGTGLDPIEAQALLATYGDPRADRLPLWGISIKSNIGHTSAAAGVAVY"
BASE COUNT 419 a 1080 c 1238 g 488 t
ORIGIN

Query Match 47.7%; Score 48.2; DB 1; Length 3225;
Best Local Similarity 67.3%; Pred. No. 0.65;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 CCGGATGCGCTCCGCGCGCTGTCATGCGCGCGGATATCTGAGACGCGCGCTGCA 60
Db 945 CCGGCTATCCGCTCGCGGATCGTCAAGCGCGCGGATCTCGACGAGCGGCTGCTGT 1004

Qy 61 TGCACMAACCCCGCGCGCTTCCGCGCGGTCATGCGCGCC 101
Db 1005 CCTCGCTATCCCGGACCGGCTGATGCGCTGCGGCGCC 1045

Search completed: November 6, 2002, 13:29:59
Job time : 593.625 secs

REFERENCE	3 (bases 1 to 996)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000)
COMMENT	This sequence is a sing

BASE COUNT ORIGIN	181 a	327 c	290 g	180 t	18 others
----------------------	-------	-------	-------	-------	-----------

RESULT 2	BF665740/c	748 bp	mRNA	linear	EST 21-DEC-2000
LOCUS	60212407F1 NIH_MGC_56 Homo sapiens	CDNA clone	IMAGE:4280754	5'	
DEFINITION	mRNA sequence.				
ACCESSION	BF665740				
VERSION	BF665740.1	GI:11939635			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
REFERENCE	1 (bases 1 to 748)				
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D.				

BASE COUNT	ORIGIN
152 a	insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
224 c	
258 g	
114 t	

Query Match	33.3%	Score 33.6;	DB 12;	Length 748;
Best Local Similarity	59.4%;	Pred. No. 31;		
Matches 57;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0;

Qy 66 GCGCAGCAGGGGATCCGACGTATGACTCGCC 101
| | | | | | | | | | | | | | | | | |
Db 464 CCGCAGAGAGCGAGAGCGGCCGAAGGAGACGCCG 429

RESULT 3	
AO846373	
LOCUS	463 bp DNA linear GSS 25-MAY-2001
DEFINITION	Leishmania major genomic library
	LMJFVL1m18c04.y1 Leishmania major genomic clone LMJFVL1m18c04 5' similar to
	contains 3. .462 VL-ch1-type-II.5 Leishmania repetitive element ;,
	DNA sequence.
ACCESSION	AO846373
VERSION	AO846373.1 GI:6051021
KEYWORDS	GSS.
SOURCE	Leishmania major.
ORGANISM	Leishmania major


```

NOTE: Vector: pZero-2 (Invitrogen); Site-1: EcoRV.
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments. Blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT      88 a      120 c      160 g      95 t
ORIGIN

```

Query Match	32.9%;	Score 33.2;	DB 17;	Length 463;
Best Local Similarity	59.6%;	Pred. No. 37;		
Matches 56; Conservative	0;	Mismatches 38;	Indels 0;	Gaps 0;

Dy 2 TCGCCGCGCACCTCGGCGCGGAGATATTTCGACCGCTGGTACGCCGGAGAGCGGCCT 61
Db 166 TGGCATCTCTTTCGACGAGCGCAGCATGACGCGCTGGTCCGTACATGAGCGGA 222

qy	62	GGCTGCGGAGCAGGGGATCGCGCACGATGGA	95
Db	226	CGCTGAGCGGGCACAAAGCTGAGACTCGCATCGA	259

RESULT 4			
BG836681.			
LOCUS	BG836681	752 bp	mRNA
DEFINITION	Zm08_06a01_A		linear
			EST 25-MAY-2001

zm08_AAFc_ECORC_Fusarium_graminearum_inoculated_corn_ear_Zea_mays
cDNA clone zm08_06a01, mRNA sequence.

ACCESSION	BG836681
VERSION	BG836681.1
ENTRYDATE	GI:14203004
DATE	

SOURCE	Zea mays.
ORGANISM	Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (pages 1 to 152)
AUTHORS Harris, L.J., Balcer

Moors, A., Hattori, J. I., Ouellet, T., Robert, L. S., Singh, J. A., Sprott, D. and Tinker, N. A.

TITLE Expressed Sequence Tags from Developing Maize Kernels Six Days after Silk Channel Inoculation with *Fusarium graminearum* Unpublished (2001)

COMMENT
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel.: (613) 759-1314
Fax: (613) 759-6566
Email: harris1j@em.agr.ca.

FEATURES	Location/Qualifiers
source	1. .752

```

/organism="Zea mays"
/cultivar="C0430"
/db_xref="taxon:4577"
/clone="zm08_06a01"
/clone_id="Zm08_AAFc_ECORC_Fusarium_graminearum_inoculate
d_corn_ear"
/tissue_type="Developing kernels (sibcrossed)"
/day_stage="10-11 days post-silk emergence"
/node=Vector: Bluescript SK+/XhoI-EcoRI, Site_1: EcoRI;
Site_2: XhoI; Field-grown maize ears were silk
Fusarium graminearum macroconidial suspension (500,000
spores/ml) and whole ears were collected and immediately
frozen in liquid nitrogen 6 days later."
BASE COUNT      173 a      220 c      201 g      143 t      15 others:

```

	ORIGIN
Query Match	33.5%; Score 32.8; DB 12; Length 753;
Best Local Similarity	58.0%; Pred. No. 50;
Matches	58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy	2	TGCGCCGGCAGCTGGGCCGGAGATATTGGACCGGTGTACGCCGGAAGAAGGGCGT	61
Dd	547	TGCGCCCTGCCTCGCCGTGGAGGGGGGTAGAGAGCGCCGCCGGTGACTTTGCCAAGTCT	608
Oy	62	GCGTCGGCAGCAGGGGATTCGCGCACCTGATGTGAACTCGCG	101
Dd	607	GCGAAATCCACCGGTGACC GGCGGAGCGATGTCAAGCGCGG	646

RESULT 5
BE360248

LOCUS	BB360248	580 bp	mRNA	linear	EST 20-JUL-2
DEFINITION	Dgl_62_F06.b1_A002 Dark Grown 1 (Dgl) sorghum bicolor cDNA, mRNA				

sequence.	
ACCESSION	BE360248
VERSION	BE360248.1
	GI:9301805

KEYWORDS	EST.
SOURCE	sorghum.
ORGANISM	Sorghum bicolor

REFERENCE

1 (bases 1 to 580)

clade; Panicoidae; Andropogoneae; Sorghum.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poaceae; PACC

AUTHORS
Cordonnier-Pratte, M.-M., Gingle, A., Maisala, C., Sudman, M. and Pratte, L.H.

TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL unpublished (2000)

COMMENT
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions

below phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 539

POLYA=NO.	Location/Qualifiers
FEATURES	1..580
SOURCE	

```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone-lib="Dark Grown 1 (DG1)"
/notes="Organ: 5-day-old dark-grown seedlings; Vector:
lambda Zap; Site.1: XhoI; Site.2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      104 a      189 c      172 g      115 t

```

ORIGIN

[illegible]

LOCUS	BQ217222	1323 bp	mRNA	linear	EST 02-MAY-2002
DEFINITION	AGNCNCURF_7569437 NIH_MGC_92		sapiens	CDNA clone	IMAGE:6043949
ACCESSION	BQ217222				
VERSION	BQ217222.1	GI:20398622			

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1323)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13285 row: m column: 06
High quality sequence start: 40
High quality sequence stop: 180.

FEATURES
source
1..1323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; Oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library." 4 others

BASE COUNT 292 a 325 c 520 g 182 t
ORIGIN

Query Match 32.3%; Score 32.6; DB 14; Length 1323;
Best Local Similarity 60.2%; Pred. No. 60;
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 13 CTCGGCGCGGAGATTTTGGACCGCTGTCAGCCGAGAGCGGGCGTGGCGCGAG 72
Db 747 CACGGCGGTGTGATGTGTCGCGCGCGCGCGTGGCGCGCGCGTGGCGCGG 806
QY 73 CAGGGGATCGCGACGTGATGACTGCC 100
Db 807 TCGAGGAGGGCCCGCGGAGAACACGC 834

RESULT 7
LOCUS BF250997 699 bp mRNA linear EST 15-NOV-2001
DEFINITION EST118254 Coccidioides immitis spherule cDNA library Coccidioides immitis cDNA clone CIAAB07 5' sequence, mRNA sequence.
ACCESSION BF250997
VERSION BF250997.1 GI:16931140
KEYWORDS EST.
SOURCE Coccidioides immitis.
ORGANISM Coccidioides immitis.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; Mitosporic Onygenales; Coccidioides.

REFERENCE 1 (bases 1 to 699)
AUTHORS Gardner, M.J., and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardneretlgr.org.

FEATURES
source
Location/Qualifiers
1..699
/organism="Coccidioides immitis"
/db_xref="taxon:5501"
/clone_lib="CIAAB07"
/dev_stage="spherule"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 136 a 206 c 186 g 171 t
ORIGIN

Query Match 32.1%; Score 32.4; DB 12; Length 699;
Best Local Similarity 68.2%; Pred. No. 63;
Matches 45; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 ATCGCCGCCACCTCGCGCGGAGATTTTTCGACCGCTGTACGCCGAGAACGGCGC 60
Db 98 ATAGCGGTCATGTAGCGCGGAGGCTTTGTACTGTCTCGGACTCTGAGAAACGGAG 157
QY 61 TGGCTG 66
Db 158 TTGATG 163

RESULT 8
LOCUS B78358 403 bp DNA linear GSS 16-JUN-1998
DEFINITION T32E17TF TAMU Arabidopsis thaliana genomic clone T32E17, DNA sequence.
ACCESSION B78358
VERSION B78358.1 GI:2774997
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 403)
AUTHORS Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Adams, M.D. and Venter, J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3
JOURNAL Unpublished (1997)
COMMENT Other_GSS: T32E17TR
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsleyetlgr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 403.

FEATURES
source
Location/Qualifiers
1..403
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="T32E17"
/clone_lib="TAMU"
/sex="hermaphrodite"
/note="Vector: pBluescript SK(-); Site_1: HindIII; Site_2: HindIII; Produced by Rod Wing"

BASE COUNT 69 a 123 c 131 g 80 t
ORIGIN

Query Match 31.9%; Score 32.2; DB 17; Length 403;
Best Local Similarity 63.6%; Pred. No. 67;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY	24	GATATTGTGGACCGTGGTAGCCGAGAAAGCGGGCCTGGCCGAGCAAGGGGATGC	83
Df	173	GAATTGGCGACTTCGCCGAMGTGGAGAGCGGCGCTTGTTGTCAGAGACCGATSAACGC	232
OY	84	GCACGTGATGACACTCGC	100
Df	233	GACATCATMTCTCACGC	249
RESULT 9 CONS05PK8 LOCUS			
DEFINITION	CNS05PK8	1050 bp	DNA linear GSS 26-MAY-2000
SOURCE	Tetraodon nigroviridis genome survey sequence T7 end of clone 006N06 of library B from Tetraodon nigroviridis, genomic survey sequence.		
VERSION	AL348065	GI:8241835	
KEYWORDS	GSS; genome survey sequence.		
ORGANISM	Tetraodon nigroviridis. Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphia; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 1050) Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher.C., Bernot,A., Fitzames,C., Winkler,P., Brottier,P., Quetier.F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence Unpublished 2 (bases 1 to 1050) Roest-Crollius,H., Jalllon,O., Dasilva,C., Fitzames,C., Fisher.C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Unpublished 3 (bases 1 to 1050) Genoscope. Direct Submission Submitted (12-APR-2000) This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon . Location/Qualifiers 1..1050 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /collection="006N06" /clone_lib="B" /note="Genoscopy sequence ID : COAB006Dc03Cl-end : T7"		
BASE COUNT	250 a	231 c	259 g 268 t 42 others
ORIGIN			
Query Match	31.9%	Score 32.2:	DB 17; Length 1050;
Best Local Similarity	57.6%;	Pred. NO.75;	
Matches	49; Conservative	0; Mismatches	36; Indels 0; Gaps 0;
OY	3	CGCCCCACCCTCGCGCGAGATTTTGGACACGCGTGTGATGGCGGAGAAGCGGGCTG	62
Df	965	CGCGCGCGCGCGCGCGGGGNGNGNGNCAGANGCTCTGCGCGCGCGCGCGCGCGCNMG	1024
OY	63	GCTGCGGAGACGAGGGGATCGCGCAC	87
Df	1025	CCNNCGCGCGCGCGCGCTCGCGC	1049
RESULT 10 BM379457/c LOCUS			
	494 bp	mRNA	linear EST 16-JAN-2002

DEFINITION	MEST505-H02.univ ISUM6 Zea mays cDNA clone MEST505-H02_3', mRNA sequence.
ACCESSION	BM379457
VERSION	BM379457.1 GI:18178247
KEYWORDS	EST.
SOURCE	Zea mays.
ORGANISM	Zea mays.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 494) Men,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S. Expressed Sequence Tags from B73 Maize: various stages and tissues including seedlings treated with a variety of hormones Unpublished (2001)
AUTHORS	Contact: Patrick S. Schnable
TITLE	Schnable Laboratory
JOURNAL	Iowa State University
COMMENT	G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA Tel.: 515-294-0975 Fax: 515-294-2299 Email: schnable@state.edu Individual basecall and confidence value were assigned using the phred software(http://www.phrap.org/). Overall sequence quality assessment and vector trimming was conducted using the Lucy software (version 1.16s, http://www.tigr.org/software/lucy). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers using a Perl program (est_process.pl), written by Dr. Hui-Hsien Chou. PCR Primers FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG) BACKWARD: primer Sp6 (ATT TAG GTG ACA CTA TAG) Seq primer: universal (GTA AAA CGA CGG CCA GT) Poly-A-tails. Location/Qualifiers 1..494 /jorganism="Zea mays" /cultivar="B73" /db_xref="taxon:4577" /cclone="MEST505-H02" /cloned_lid="ISUM6" /tissue_type="mixed" /lab_host="DH10B" /note=Vector: pSP1p7 (4.43 kb); Site_1: EcoRI; Site_2: Note: Tissue samples were collected and partially pooled prior to RNA extraction. First-strand cDNAs were prepared from 21 individual pools of oligo-dT selected mRNAs by priming with 21 different NotI oligo-dT tag primers (5'-AACGTGAGAGAATTCGGCGCCGCNNNNTTTTTTTTTTTTTT-3'). Distinguishable 'bar code' tags, (N)6, were used for each separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are: ATMGCG--Germinated seeds and seedlings (1, 2, 8, 11 DAG); ACTGCC--Mixed mature tissues (17, 21, 39, 69, 77 DAG); CACGCC--Kernels (3, 5, 10, 15, 20, 25, 30, DAG); TAACGC--Adventitious roots (65 DAG); CAGCGC-- tassels (3-39 cm, 53 and 56 DAG); AGCTAC--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG); TGAGCG--Husks (73 DAG); GACCAC--Silks; AMTGCG--unpollinated first ears; CTAGCG--ear shanks; GAAGAG--etiolated seedlings; AGTAGC--callus; GTGCAC--Cycloheximide-treated callus; GTGACC--Anaerobic treated seedlings; GGTCCA--NA (α-Naphthalene acetic acid)-treated seedlings; GATGCC--Kinetin-treated seedlings; AAGCGC--ACPC (l-amincyclop propane-l-carboxylic acid-seedlings; CTAGCC--ABA (Abscicic acid)-treated seedlings; TAOGGA-GA (gibberellic acid)-treated seedlings; GAGGA--JA (jasmonic acid)-treated seedlings; Equal

amounts of first-strand cDNA from each reaction were combined and used as template for DNA PolI-catalyzed 2nd strand synthesis. After the addition of EcoRI adaptors, ds-cDNAs were digested with NotI. Molecules between 0.5 and 2.0 kb were directionally cloned into the EcoRI and NotI sites of the pSLIP7 expression vector. Plasmid DNA isolated from the library was digested with NotI to remove empty vector clones. Linear DNAs from 5.4 to 7 kb were gel purified and ligated at low concentration to promote recircularization. Ligation products were precipitated and transformed into DH10b host cells. The complexity of the resulting library was 1.2 x 10⁶."

BASE COUNT 118 a 136 c 143 g 97 t
ORIGIN

Query Match 31.5%; Score 31.8; DB 13; Length 494;

Best Local Similarity 64.0%; Pred. No. 88;
Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 2 TCGCCCGCAGCTCGCGGAGATATTTCGACCGCTGTACCGCGAGAACGGCGCT 61

DB 479 TCGCCAGCTCCGCCCGCGGAGACTTCTCGCGGCTGTCTGCTTCCCTGCACGCGCTCGCG 420

OY 62 GCGTGGCGAGCAGG 76

DB 419 TCGCGCGCTGGAGG 405

RESULT 11

AY108053

LOCUS Zea mays PCO136645 mRNA sequence. 558 bp mRNA linear HTC 25-MAY-2002

ACCESSION AY108053

VERSION AY108053.1 GI:21211131

KEYWORDS HTC.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 558)

AUTHORS Halney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,

Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

Maize Mapping Project/Dupont Consensus Sequences for Design of

Overgo Probes

Unpublished (2002)

2 (bases 1 to 558)

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission

Submitted (25-Apr-2002) Maize Mapping Project, University of

Missouri, Columbia, MO 65211, USA

FEATURES

source

1. 558

/organism="Zea mays"

/db_xref="MaizeDB:638015"

/db_xref="taxon:4577"

/clone="PCO136645"

/clone_1lb="Maize Mapping Project/Dupont Consensus

library"

/note="this sequence is part of a project of EST

assemblies resulting from the application of public

contigs to seed Dupont contigs; this resource was

assembled by Dupont as part of a collaboration for the

overgo addressing of BACs in conjunction with the Maize

Mapping Project"

BASE COUNT 112 a 164 c 153 g 129 t

ORIGIN

Query Match 31.5%; Score 31.8; DB 11; Length 558;

Best Local Similarity 64.0%; Pred. No. 89;

Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 2 TCGCCCGCAGCTCGCGGAGATATTTCGACCGCTGTACCGCGAGAACGGCGCT 61

DB 81 TCGCCAGCTCCGCCCGCGGAGACTTCTCGCGGCTGTCTGCTTCCCTGCACGCGCTCGCG 140

OY 62 GCGTGGCGAGCAGG 76

DB 141 TCGCGCGCTGGAGG 155

RESULT 12

A0655342/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Other GSSs: Sheared DNA-1014.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared

DNA library constructed at TIGR. Clones will be available for

distribution through ATCC. Sheared DNA end sequences search page:

<http://www.tigr.org/ldb/ldb/seq/>

Seq primer: M13-Reverse

Class: shotgun.

Location/Qualifiers

1. 573

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db_xref="taxon:5691"

/clone="Sheared DNA-1014"

/clone_1lb="Sheared DNA"

/note="Vector: pUC18; Site1: Sma1; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically

sheared to give a tight size distribution (approx 2 kb).

The v + i method used for the library construction is

described in detail in Smith, H.O. and Venter, J.C.

(Making small insert libraries for whole genome shotgun

sequencing projects. In Genome Sequencing: A Practical

Approach, eds. M. Vaudin and B. Barrell, Oxford University

Press, 1999).

BASE COUNT 118 a 172 c 152 g 131 t

ORIGIN

Query Match 31.5%; Score 31.8; DB 17; Length 573;

Best Local Similarity 64.0%; Pred. No. 89;

Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 26 TATTTCGACCGCTGTACCGCGGAGAGCGGCGTGTCCGCGAGCAGGAGATCCGC 85

DB 489 TATTTCGCTCCGCTGTACCGCGGAGAGCGGCGTGTCCGCGAGCAGTGTGACCTT 430

OY 86 AGCTGATGACATCGC 100

DB 429 GCGTGGAGGATATGC 415

[illegible]

ACCESSION	Sheared DNA-34021, DNA sequence.		
VERSION	AQ940375		
KEYWORDS	AQ940375.1 GI:6763640		
SOURCE	GSS.		
ORGANISM	Trypanosoma brucei.		
	Trypanosoma brucei		
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
REFERENCE	1 (bases 1 to 640)		
AUTHORS	El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ollu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.		
TITLE	Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library		
JOURNAL	Unpublished (1999)		
COMMENT	Other_GSSs: Sheared DNA-34021.TF Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/ . Seq primer: M13-Reverse Class: shotgun.		
FEATURES	location/Qualifiers		
SOURCE	1. 640		
	/organism="Trypanosoma brucei"		
	/strain="TREU927/4 GUTat 10.1"		
	/db_xref="taxon:5691"		
	/clone="Sheared DNA-34021"		
	/clone.lib="Sheared DNA"		
	/note="vector: pUC18; Site.1: SmaI: Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."		
BASE COUNT	150 a 168 c 190 g 132 t		
ORIGIN			
Query Match	31.5%, Score 31.8; DB 17; Length 640;		
Best Local Similarity	64.0%, Pred. No. 90;		
Matches	48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;		
OY	26	TATTTCGACCGCTGTACGCCGAGAGCGGGCGTGCCTGCGGACGAGGGGATCGCGC	85
		111 11 11 111 111	
Db	199	TATTCTGTCCGTCGTACGCGCGGACGCGTGCAGCAGCAGTGAGACCTT	258
OY	86	ACGTGATGAGACTCGC	100
		111	
Db	259	GCGTGGAGGATATGC	273
RESULT 15			
BH839856/c			
LOCUS	BH839856	737 bp	DNA linear GSS 28-MAY-2002
DEFINITION	LMCR200005F01f Zea mays L. Zea mays genomic clone LMCR200005F01f, DNA sequence.		
ACCESSION	BH839856		
VERSION	BH839856.1 GI:21237924		
KEYWORDS	GSS.		
SOURCE	Zea mays.		
ORGANISM	Zea mays		

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 737)
 Kim, S.W., Yu, Y., Lee, M.C., Main, D. and Wang, R.A.
 Methyl-filtration genomic sequence from maize
 Unpublished (2002)
 Contact: Wang RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Total High Quality bases = 402
 Seq primer: TAAATAGACTACTATATAGGC
 Class: shotgun
 High quality sequence start: 10
 High quality sequence stop: 520.
 Location/Qualifiers

FEATURES

source
 1..737
 /organism="Zea mays"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="LMCR200005F01f"
 /clone_lib="Zea mays L."
 /issue_type="Leaf"
 /lab_host="DH10B"
 /note="vector: pGEM-T easy; Site 1: Mcr BC;
 methyl-filtration library; Nuclei DNA was completely
 digested with Mcr BC, size fractionated and transformed
 to E.Coli.DH10B."

BASE COUNT 117 a 249 c 210 g 158 t 3 others
 ORIGIN

Query Match 31.5%; Score 31.8; DB 17; Length 737;
 Best Local Similarity 61.4%; Pred.No.92;
 Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 19 GCGAGATATTTCGACCGCTGTGACCGCGAGACCGCGGTGCGTGCAGACAGGG 78
 ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 DB 723 GCGAGCTTTACGCGGACCGCGAAGCCGAGAGCGGTAGTGGAAGGAGAGAGT 664
 ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 79 ATGCGCACGCTGATGACTGCG 101
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 DB 663 GTAGAGCGCGGTTGGCTTGCG 641

Search completed: November 6, 2002, 15:52:44
 Job time : 1201.62 secs

Query Match 32.7%; Score 33; DB 10; Length 594;
Best Local Similarity 63.0%; Pred. No. 0.18;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 15 CGCGCGGAGATTTTCCAGCCCTGTACCGCGGAGAGAGCGCGCTGGCTGGCGGACA 74
DB 222 CAGCCCGGAGTCTCGCGCCGCGGAGGCGGAGGCGGCTTGGCTGGCTT 163

QY 75 GGGATCGCGCAGCTGATGGA 95
DB 162 GAGGATGCGCTTCTCGACGAA 142

RESULT 2
US-09-832-320-3
; Sequence 3, Application US/09832320
; Patent No. US20010049834A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Pathogenesis-Related
; FILE REFERENCE: 35718/214291
; CURRENT APPLICATION NUMBER: US/09/832,320
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/195,801
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(612)
US-09-832-320-3

Query Match 28.9%; Score 29.2; DB 10; Length 612;
Best Local Similarity 65.2%; Pred. No. 2.3;
Matches 43; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 31 GCGACCGCTGTACGCCGGAAGCGGCGTGGCTGGCGGAGGAGGATCGCGACGCTG 90
DB 283 GCGCGCGCGGACGCTGGCGGACGAGCGGCGGAGGCGGCGGCTTCCGCGGACGCTG 342

QY 91 ATGGAC 96
DB 343 GGGGCC 348

RESULT 3
US-09-832-320-1
; Sequence 1, Application US/09832320
; Patent No. US20010049834A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Pathogenesis-Related
; FILE REFERENCE: 35718/214291
; CURRENT APPLICATION NUMBER: US/09/832,320
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 60/195,801
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:

NAME/KEY: CDS
LOCATION: (63)...(674)
US-09-832-320-1

Query Match 28.9%; Score 29.2; DB 10; Length 898;
Best Local Similarity 65.2%; Pred. No. 2.3;
Matches 43; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 31 GCGACCGCTGTACGCCGGAAGCGGCGTGGCTGGCGGAGGAGGATCGCGACGCTG 90
DB 345 GCGCGCGCGGAGCGGCTGGCGGACGAGCGGCGGAGGCGGCGGCTTCCGCGGACGCTG 404

QY 91 ATGGAC 96
DB 405 GGGGCC 410

RESULT 4
US-09-741-669-159
; Sequence 159, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1518)
US-09-741-669-159

Query Match 28.7%; Score 29; DB 10; Length 1518;
Best Local Similarity 71.7%; Pred. No. 2.7;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 49 GAGAGCGGCGCTGTGCTGGCGGAGGAGGATCGCGACGATGATGACTCGG 101
DB 70 GAGAGCTGGCGAACCCTGGCGGAGGAGGATTCCTCCGAAAGATTTCG 122

RESULT 5
US-09-815-242-6188
; Sequence 6188, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6188
;; LENGTH: 1518
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1518)
US-09-815-242-6188

Query Match 28.7%; Score 29; DB 10; Length 1518;
Best Local Similarity 71.7%; Pred. No. 2.7;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 49 GAGAACGGCGGTGGCTGCGAGAGGAGGATCGCGACGTGATGACTCGCG 101
DB 70 GAGAGCTGGCGAACCTCGCGAGCGAGGATGCTTCGCCGACGATTTCCG 122

RESULT 6
US-09-815-242-7886

;; Sequence 7886, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyakind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7886
;; LENGTH: 1056
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1056)
US-09-815-242-7886

Query Match 27.1%; Score 27.4; DB 10; Length 1056;
Best Local Similarity 57.6%; Pred. No. 7.6;
Matches 49; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 ATCGCCCGCCACCTCGCGCGAGATATTGGACCGCTGTACCGCGGAGAGCGGCG 60
DB 718 ACCGCGTTGCTCTGCTGTCGAGTCTCGCCAGCGCGCTACCGGAGCGCTGAAG 777
OY 61 TGGCTCGCGGACGAGGATCGCGC 85
DB 778 GGGCCGCGCGCGCGCGCGCACCGCGC 802

RESULT 7
US-09-764-847-1301

;; Sequence 1301, Application US/09764847
;; Patent No. US20020132767A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009
;; CURRENT APPLICATION NUMBER: US/09/764,847
;; CURRENT FILING DATE: 2001-01-17
;; PRIOR APPLICATION data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2003
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1301
;; LENGTH: 15987
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-847-1301

Query Match 26.3%; Score 26.6; DB 10; Length 15987;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 50; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 7 CGCCACCTCGCGCGGAGATATTGCGACCGCTGTACCGCGGAGAGCGGCGCTG 66
DB 615 CGGCTCGCGGAGAGAGCGTCCGGCGCGGAGAGTGGTCTGGGGCTTGGCG 674
OY 67 CGCGAGAGGAGATCGCGCACGTGATGGA 95
DB 675 GCGGAGTAGTAGGAGCGCGGAGGCGCTGGA 703

RESULT 8

US-09-960-352-1020/c
;; Sequence 1020, Application US/09960352
;; Patent No. US20020137139A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Tao, Nengding
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathialagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 1020
;; LENGTH: 314
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; OTHER INFORMATION: Clone ID: 05-LIB3058-001-Q1-K1-B1
US-09-960-352-1020

Query Match 26.1%; Score 26.4; DB 10; Length 314;
Best Local Similarity 57.1%; Pred. No. 14;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 5 CCGCGCACCTCGCGCGGAGATATTGGACCGCTGTACCGCGGAGAGCGGCGTGGC 64

Db	139	CTGGGCAACCCACACTTGGCAGTTCTCGGTGTCCCAAGCTGCAGGGCCGAAGGGTCCGGCA	80
QY	65	TGCGCAGCAGGGGATCGGCACG	88
Db	79	CAGCGGCGCGGGGGGTGGCGCACG	56

RESULT 9
US-10-044-090-842/c
; Sequence 842, Application US/10044090
Data+No 10000201270811

```

? TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
? FILE REFERENCE: PA-0028 US
? CURRENT APPLICATION NUMBER: US/10/044,090
? CURRENT FILING DATE: 2002-01-09
? NUMBER OF SEQ ID NOS: 850
? SOFTWARE: PERL Program
? SEQ ID NO 842
? LENGTH: 3172
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: incyte ID No. US20020137081A1 410580.16
? NAME/KEY: unsure
? LOCATION: 2836, 2838
? OTHER INFORMATION: a, t, c, g, or other
? US-10-044-090-842

```

Query Match	26.1%;	Score 26.4;	DB 12;	Length 3172;
Best Local Similarity	.61.8%;	Pred. No. 15;		
Matches	42;	Conservative	0;	Mismatches 26;
				Indels 0;

Oy 12 CCTGGCGCCGAGATATTGCCACCCCTGTGTCAGCCGGACAAGCGGCCTGGCTGGCGGA 71
 | | | | | | | | | | | | | | | |
Db 817 CATGGGCACAGACACTGCTCCGCTGTGATTCTGGAAGAGGCGCGGTGGGTGCTGGGA 75B

QY	72	GCAGGGGA	79
Db	757	GTCGGGGA	750

RESULT 10
US-09-880-107-2097/c
; Sequence 2097, Application US/09880107
; Patent No. US20020142981A1
; Inventor: TUNNEY, JAMES

APPLICANT: Vockley, Joseph G.

APPLICANT: Gene Logic, Inc.

FILE REFERENCE: 44921-5028-WO

COCKENI FILING DATE: 2001-06-14
 PRIORITY NUMBER: 00 60 4311 370

PRIOR FILING DATE: 2000-06-14

PRIOR FILING DATE: 2000-10-02

SOFWANE: PALENCIN VEL. 2.1
GEO TD NO 2007

GENEAL. 1/2003
TYPE: DNA

ORGANISM. HOMO SAPIENS.

OTHER INFORMATION: GENIDALIAK ACCESSION NO. 0520020142361A1 003/04
US-00-880-107-2007

Query Match 26.18; Score 26.4; DB 10; Length 17509

Best Local Similarity 61.8%; Pred. No. 16;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 12 CCTCGCGCGGAGATATTTCGACCGCTGTGTACCGCGGAGAAGCGGCGTGTGCTGGCGCA 71
| | | | | | | | | | | | | | | | | | | | | |
Db 8113 CATGGGCACAGACAGTGCCTGCCGCTGATTTGTGGAAGAGCGCGCGTGTGCTGCA 8054

QY	72	GCAGGGGA	79
Db	8053	GTCGGGGA	8046

RESULT 11
US-09-878-574-4662
; Sequence 4662, Application US/09878574
; Patent No. US20020110548A1
; Generalization

```

: APPLICANT: Byrum, Joseph R.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Thompson, Michael D.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(15401)B
: CURRENT APPLICATION NUMBER: US/09/878,574
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 09/733,535
: PRIOR FILING DATE: 1999-06-14
: NUMBER OF SEQ ID NOS: 15775
: SEQ ID NO 4662

```

```

: ORGANISM: Glycine max
:
: FEATURE:
:
: NAME/KEY: unsure
:
: LOCATION: (1)..(785)
:
: OTHER INFORMATION: unsure at all n locations
:
: OTHER INFORMATION: Clone ID: LIB3028-039-Q1-B1-HB
:
: OS-09-878-574-4662

```

Query Match	25.98;	Score 26.2;	DB 10;	Length 785;
Best Local Similarity	58.28;	Pred. No. 17;		
Matches 46;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;

Oy 16 GGCCCGAGATATTTCGACGCCGTGTATGCCGGAGAAGCGGGCTGGCTGCAGCAG 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 GGCAGGGAGGTACAGACAGGGTGGGGCCCGGTTGGGAGAGAGCGCGGGCGGACAG 434

QY	76	GGGATCGCGCAGTGATGG	94
Db	435	GCGGTGCGCAAGCGAAGG	453

RESULT 12
US-09-925-302-47
; Sequence 47, Application US/09925302
; Patent No. US20020044941A1

APPLICANT: Rosen et al

FILE REFERENCE: PA104

CURRENT FILING DATE: 2001-08-10

PRIOR FILING DATE: 2000-03-08

PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ. ID NOS.: 806

SOFTWARE: Pallenlii Ver. 2.0
CDO TO NO 17

LENGTH: 3/13

US-09-925-302-47

```
Query Match          25.9%; Score 26.2; DB 10; Length 3773;
Best Local Similarity 56.3%; Pred. No. 17;
Matches 49; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 8 GCCACCTCGCGCGAGATATTTCGACCGCGAGAGCGGCGCTGC 67
    ||| | | | | | | | | | | | | | | | | | | | | |
Db 952 GCCAGCGGACGAGAGAGGTCGCCGCTGACACAGCTGGAGAAACAGACGCG 1011
QY 68 GCGAGCAGGAGGATCGCGACGTGATGG 94
    ||| | | | | | | | | | | | | | | | | | | |
Db 1012 TGCAGCAGAGCTGTGACGACCTGCTGG 1038

RESULT 13
US-09-878-574-11248
; Sequence 11248, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878, 574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 11248
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701064227H1
US-09-878-574-11248

Query Match          25.7%; Score 26; DB 10; Length 259;
Best Local Similarity 55.6%; Pred. No. 19;
Matches 50; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 6 CCGCCACCTCGCGCGAGATATTTCGACCGCTGTGACGCGGAGAGCGGCGTGGCT 65
    ||| | | | | | | | | | | | | | | | | | | | | |
Db 50 CCAGGACCTCGCTGCGCGGATGACACGCGCGGCTGGGAGAGAGCGCGCAT 109
QY 66 CCGCGAGCAGGAGATCGCGCAGTGATGA 95
    ||| | | | | | | | | | | | | | | | | | | |
Db 110 GGAGAGCTCTGCATCGACGACATGCTGA 139

RESULT 14
US-09-983-965-5707
; Sequence 5707, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983, 965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465, 231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113, 678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5707
; LENGTH: 316
; TYPE: DNA
; ORGANISM: Bos taurus
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: 56-L1B34-038-Q1-E1-F12
US-09-983-965-5707

Query Match          25.7%; Score 26; DB 10; Length 316;
Best Local Similarity 57.3%; Pred. No. 19;
Matches 47; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 17 GCGCGAGATATTTCGACCGCTGTGACCGCGAGAGCGGCGCTGCCTGCCGAGCAGG 76
    ||| | | | | | | | | | | | | | | | | | | | | |
Db 172 GCGCAGACGCTTTGTGTGACTGAACGTGAGCTGAGTGAGTGAGTGATGCGAGGTGT 231
QY 77 GGATCGCGACGCTGATGACTC 98
    ||| | | | | | | | | | | | | | | | | | | |
Db 232 CTGTCTCATGTGATGTGCTCC 253

RESULT 15
US-09-822-830A-286/c
; Sequence 286, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822, 830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 286
; LENGTH: 1601
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1490
; OTHER INFORMATION: n=a,c,g, or t
US-09-822-830A-286

Query Match          25.7%; Score 26; DB 10; Length 1601;
Best Local Similarity 58.7%; Pred. No. 19;
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 4 GCCCGCACCTCGCGCGAGATATTTCGACCGCTGTGACCGCGAGAGCGGCGCTGG 63
    ||| | | | | | | | | | | | | | | | | | | | | |
Db 1495 GGCTGNCCTCTGCGCGCGGAGATGCTGGACCGACGCTGGCGAGAGAGCGCGCTGC 1436
QY 64 CTGCGCAGCAGGAGG 78
    ||| | | | | | | | | | | | | | | | | | | |
Db 1435 CCGCTGAGAGGAGG 1421

Search completed: November 6, 2002, 20:32:52
Job time : 34.125 secs
```



```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 41821
OY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 41821
OY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 41821
OY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 60
DB 41762 ATGCCCCGACCTCGCGCGGAGATATTGGACCGCTGTACGCCGGAAGCGGGCG 41821
OY 61 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 101
DB 41822 TGGCTGCGGAGCAGGAGATCGCGCACGTGATGACTCGCG 41862
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 60
DB 41762 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 41821

OY 61 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 101
DB 41822 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 41862

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 60
DB 41762 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 41821

OY 61 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 101
DB 41822 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 41862
```

```
DB 41822 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 41862

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 60
DB 41762 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 41821

OY 61 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 101
DB 41822 TGGCTGCGCGAGACGAGGATGCGCAGCTGATGACTCGCG 41862

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          43.0%; Score 43.4; DB 4; Length 4403765;
Best Local Similarity 64.4%; Pred. No. 0.00079;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 ATCGCCCGCCACCTGCGGCGGAGATATTGCGACCGCTGTGACCGCGAGAGCGGGCG 60
```

Db 4287107 ATAGCCGCGCGCGGCGGCTAGATCTTCGCCACCGCGGTACCCACAGCGTGGAGAG 4287048
OY 61 TGGCTGGCGAGCAGGGGATCGCGACGATGATGACTCGCG 101
Db 4287047 CTAATGCGCACATGGGTATGAGCATGTCTATGACTCGCG 4287007

RESULT 10

US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 43.0%; Score 43.4; DB 4; Length 4411529;
Best Local Similarity 64.4%; Pred. No. 0.00079;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 ATGCCCCGCACTCGCGCGGAGATTTGCGACCGCTGTACGCCGGAAGCGGCG 60
Db 4294850 ATAGCCGCGCGCGGCGGCTAGATCTTCGCCACCGCGGTACCCACAGCGTGGAGAG 4294791
OY 61 TGGCTGGCGAGCAGGGGATCGCGACGATGATGACTCGCG 101
Db 4294790 CTAATGCGCACATGGGTATGAGCATGTCTATGACTCGCG 4294750

RESULT 11

US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 2 TCGCCCCGCACTCGCGCGGAGATTTGCGACCGCTGTACGCCGGAAGCGGCG 59
Db 12950 TCGCACGCGCACCTCGGCGCGGAGTCTTTCGCCACCGCATCCAGGGAAGTGGAGCGC 13007

RESULT 12

US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-456-837-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 2 TCGCCCGCCACTCGGCGCGAGATATTTCGACCGCTGGTACCGCCGAGAGCGGCG 59
Db 12950 TCGCAGCCACCTCGGCGCGAGGTCTTTGCGACCGCCAGTCAGAGGTGAGCGC 13007

RESULT 13
US-08-457-342-6
Sequence 6, Application US/08457342
Patent No. 5662898
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,342
FILING DATE: 01-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-342-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 2 TCGCCCGCCACTCGGCGCGAGATATTTCGACCGCTGGTACCGCCGAGAGCGGCG 59
Db 12950 TCGCAGCCACCTCGGCGCGAGGTCTTTGCGACCGCCAGTCAGAGGTGAGCGC 13007

RESULT 14
US-08-457-646A-6
Sequence 6, Application US/08457646A
Patent No. 5679560
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,646A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-646A-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Search completed: November 6, 2002, 17:26:15
Job time: 1793.62 secs

Db 12950 TCGCAGCCACCTCGGCGCGAGTCTTTCACCGCCAGTCCAGGAGTGGAGCGC 13007

RESULT 15

US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Lyon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-458-076A-6

Query Match 32.1%; Score 32.4; DB 1; Length 28958;
Best Local Similarity 72.4%; Pred. No. 0.72;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 12950 TCGCAGCCACCTCGGCGCGAGTCTTTCACCGCCAGTCCAGGAGTGGAGCGC 13007

misc_rna 2031..3548

```

FT      /tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /tag= x

FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /tag= ad
FT      /label= epob_gene
FT      /note= "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /tag= ao
FT      /note= "encodes DHS"
FT      misc_RNA
FT      29433..30287
FT      /tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /tag= au

```

```
FT /note= "encodes DH6"
FT misc_RNA
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS
FT /tag= ay
FT /label= "epoE_gene"
FT /note= "encodes modules 7 and 8"
FT misc_RNA
FT /tag= az
FT /note= "encodes KS7"
FT misc_RNA
FT /tag= ba
FT /note= "encodes AT7"
FT misc_RNA
FT /tag= db
FT /note= "encodes KR7"
FT misc_RNA
FT 42181..42423

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. NO. 8.4e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGCCGCCACCTCGCGCGAGATATTGGACCGCTGTGACCGCGAGAGAGCGGCG 60
Db 36150 ATCGCCGCCACCTCGCGCGAGATATTGGACCGCTGTGACCGCGAGAGAGCGGCG 36209
QY 61 TGGCTGCCGCGAGCAGCGGATCGCGACGTGATGACTCGCG 101
Db 36210 TGGCTGCCGCGAGCAGCGGATCGCGACGTGATGACTCGCG 36250

RESULT 2
AA255887
ID AA255887 standard; DNA; 68750 BP.
XX
AC AA255887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Location/Qualifiers
FT 1..1826
FT /tag= a
FT /partial=
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS
FT /tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT complement (5612..5992)
FT /tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /tag= f
```

```
FT /product= "Type I polyketide synthase, EPOS A"
FT 11872..116104
FT /tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P"
FT 16251..21749
FT /tag= h
FT /product= "Type I polyketide synthase, EPOS B"
FT 21746..43519
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C"
FT 43524..54920
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D"
FT 54935..62254
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E"
FT 62369..63628
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F"
FT 63779..64333
FT /tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT complement (63853..64290)
FT /tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT complement (64287..64727)
FT /tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /tag= q
FT /product= "Orf 10 protein"
FT complement (65008..65874)
FT /tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT complement (65871..66338)
FT /tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT 67334..68251
FT /tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT 68346..68750
FT /partial=
FT /tag= v
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
FT W09966028-A2.
FT 23-DEC-1999.
XX
PD 16-JUN-1999; 99WO-EP04171.
XX
PE 18-JUN-1998; 98US-0099504.
XX 24-SEP-1998; 98US-0101631.
PR 05-FEB-1999; 99US-0118906.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

XX	WPI: 2000-097741/08.
DR	P-PSDB: AAY58573, AAY58574, AAY58575, AA58576, AA58577, AA58578,
DR	AAY58579, AAY58580, AAY58581, AA58582, AA58583, AA58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR	AAY58592, AAY58593, AAY58594.
XX	
PT	New isolated epoethyllone synthase genes, used for the recombinant
PT	production of epoethyllone for use in cancer therapy -
XX	
PS	Claim 14; Page 87-104; 174pp; English.
XX	
CC	This sequence represents a 68_75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethyllones. Epoethyllones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC	EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC	the thiazole ring formation of epoethyllones, and EPOS B, EPOS C, EPOS D
CC	and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC	formation. EPO F (AAY58579) is an epoethyllone macrolactone oxidase, and
CC	the proteins Orf 3 (AAY58582) and Orf14 (AA58593) are thought to be
CC	involved in transport. Epoethyllones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethyllones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficiently exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethyllones as anticancer agents, they are problematical to produce on a
CC	large scale. Epoethyllones are too complex for industrial scale chemical
CC	synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC	poor yields of epoethyllones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethyllones in a heterologous host
CC	that is more amenable to fermentation.
XX	
SQ	Sequence 68750 BP: 9596 A; 22458 C; 25537 G; 11159 T; 0 other:
	Query Match 98.4%; Score 99.4; DB 21; Length 68750;
	Best Local Similarity 99.0%; Pred. No. 2,3e-18;
	Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0
OY	1 ATCGCCGCGCACCTTCGGCGCGAGATATTGTGCACGCCGTGTAACGCCGGAAGCGGGC 60
Db	41762 ATCGCCGCGCACCTTCGGCGCGAGATATTGTGCACGCCGTGTAACGCCGGAAGCGGGC 41821
OY	61 TGCGTGGCGGACGACGAGGATCGCGCACGTGATGACTCGCG 101
Db	41822 TGCGTGGCGGACGACGAGGATCGCGCACGTGATGACTCGCG 41862
	RESULT 3
ID	AAF90037 standard; DNA; 6462 BP.
XX	AAF90037;
DT	06-AUG-2001 (first entry)
XX	
DE	Nucleotide sequence of a type I polyketide synthase.
XX	
KM	Metabolic pathway operon: polyketide; polyketide antibiotic;
XX	type I polyketide synthase; ss.
OS	unidentified.
FT	Key Location/Qualifiers
FT	CDS 1..6462
FT	/tag-a
FT	/product-"type I polyketide synthase"
FT	/transl_except="(1..3, aa: Met)"

PN	XX		MO200140497-A2.
XX	FD	07-JUN-2001.	
XX	PF	27-NOV-2000; 2000WO-FR03311.	
XX	PR	29-NOV-1999; 99FR-0015032.	
XX	PR	07-JUN-2000; 2000US-0209800.	
XX	PA	(AVET) AVENTIS PHARMA SA.	
XX	PI	Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;	
XX	PI	Cappellano C, Francou F, Reynal A, Ball M, Sezonov G, Tuphile K;	
XX	PI	Frostegard A;	
DR	WPJ:	2001-374849/39.	
XX	P-PSDB:	AAB83974.	
PT		Collection of nucleic acids from environmental samples, useful for	
PT		identifying e.g. genes encoding polyketide synthases and derived	
PT		antibiotics -	
PS		Claim 35; Page 307-309; 356pp; French.	
XX		The specification describes a method for the preparation of a collection	
CC		of nucleic acids from organisms in a soil sample. The method comprises	
CC		milling a dried sample to produce microparticles; suspending these in	
CC		liquid buffer; extraction of nucleic acids from the microparticle;	
CC		passing nucleic acid-containing solution through a molecular sieve;	
CC		passing nucleic acid-enriched fractions through an anion exchange	
CC		chromatography material; and recovering fractions containing purified	
CC		nucleic acids. The nucleic acids are sources for sequences that encode	
CC		either operons involved in a metabolic pathway (specifically polyketide	
CC		synthesis) or polypeptides, particularly for production of therapeutic	
CC		or agricultural compounds, especially polyketide antibiotics. AAF90034-39	
CC		represent open reading frames (ORFs) of the coding strand of cosmid	
CC		a26g1, and encode type I polyketide synthases.	
XX	SQ	Sequence 6462 BP; 1073 A; 2067 C; 2135 G; 1187 T; 0 other;	
		Query Match	52.5%; Score 53; DB 22; Length 6462;
		Best Local Similarity	70.3%; Pred No. 1.1e-05;
		Matches 71; Conservative	0; Mismatches 30; Indels 0; Gaps 0;
OY	1	ATCGCCGCCACCTGCGCGAGATATTTCGCACCGCTGTACGCCGGAGAAGCGGCG 60	
Db	4750	ATCGCACACGCGCGCGCGCGGAGATCTTCCGTACTCGCCGGAGATCCGAAAAACGAGCG 4809	
OY	61	TGCGTGCGCAGACAGCGGATCGCGCATGTGACTGCGG 101	
Db	4810	TATCTGCGCTCGCTGCGCATCGCGCATGTTCGGATTCGCG 4850	
RESULT 4			
AAFA90033	ID	AAF90033 standard; DNA: 34071 BP.	
XX	AC	AAF90033;	
XX	D7	06-AUG-2001 (first entry)	
XX	DE	Nucleotide sequence of cosmids a26g1 (coding strand).	
XX	KM	Metabolic pathway operon; polyketide; polyketide antibiotic; ss.	
XX	OS	Synthetic.	
XX	PN	MO200140497-A2.	
XX	PD	07-JUN-2001.	
XX	PE	27-NOV-2000; 2000WO-FR03311.	

```
XX 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics
XX
XX Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthesis) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. The present
XX sequence represents cosmid a26g1 coding strand, which encodes different
XX polyketide synthases.
XX
XX Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
XX
XX Query Match 52.5%; Score 53; DB 22; Length 34071;
XX Best Local Similarity 70.3%; Pred. No. 1.2e-05;
XX Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
XX
XX 1 ATCGCGCGCACCTCGGCGGAGATATTTCGACCGCTGGTAGCGCGGAGAGCGGCG 60
XX Db 22973 ATCGCACAGCGTCGGCGGCGGAGATCTTCCTACTGCGGAGATCCGAAAAACGAGCG 23032
XX
XX 61 TGGCTGCGCAGACAGGGGATCGCGACGTGATGACTCGCG 101
XX Db 23033 TATCTGCGCTGCGTGGCGATCGCGCATGTTCGATTCGCG 23073
XX
XX RESULT 5
XX AAF90032/C
XX ID AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX Synthetic.
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR03311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX PR 07-JUN-2000; 2000US-0209800.
XX
XX (AVET ) AVENTIS PHARMA SA.
XX
XX Jeannin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
```

```
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics
XX
XX Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthesis) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. The present
XX sequence represents cosmid a26g1 (non-coding strand). The sense strand
XX encodes different polyketide synthases.
XX
XX Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
XX
XX Query Match 52.5%; Score 53; DB 22; Length 42717;
XX Best Local Similarity 70.3%; Pred. No. 1.2e-05;
XX Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
XX
XX 1 ATCGCGCGCACCTCGGCGGAGATATTTCGACCGCTGGTAGCGCGGAGAGCGGCG 60
XX Db 11147 ATCGCACAGCGTCGGCGGCGGAGATCTTCCTACTGCGGAGATCCGAAAAACGAGCG 11088
XX
XX 61 TGGCTGCGCAGACAGGGGATCGCGACGTGATGACTCGCG 101
XX Db 11087 TATCTGCGCTGCGTGGCGATCGCGCATGTTCGATTCGCG 11047
XX
XX RESULT 6
XX AAF31641
XX ID AAF31641 standard; DNA; 6333 BP.
XX
XX AAF31641;
XX
XX 09-APR-2001 (first entry)
XX
XX Mycobacterium tuberculosis mas gene.
XX
XX Mycobacterium tuberculosis mas gene.
XX
XX Mycobacterium tuberculosis; attenuated microorganism;
XX signature tagged transposon mutant; mutant library;
XX Mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
XX vaccine; mas; mycocerosic acid synthase; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200102555-A1.
XX
XX 11-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-IB00950.
XX
XX 06-JUL-1999; 99US-0142982.
XX PR 08-JUL-1999; 99US-0142833.
XX
XX (INSP ) INST PASTEUR.
XX
XX Gicquel B, Guilhot C, Camacho L;
XX
XX WPI: 2001-091804/10.
XX DR P-PSDB; AAB66471.
XX
```


CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AA199683) and
CC H37RV (AA199682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.
XX
SO Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other:
Query Match 43.0%; Score 43.4; DB 22; Length 4411529;
Best Local Similarity 64.4%; Pred. No. 0.0062;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 1 ATGCGCCGACCTGCGCGGAGATATTGCGACCGCTGTACCGCGAGAGCGGCG 60
DB 4294850 ATAGCCCGCGCGGCGGCGTGCATCTGCGCCACCGCGTACCCGACGCGTGCAGAG 4294791
OY 61 TGGCGCGGAGAGGAGGATGCGCGAGCTGATGACCTGCGG 101
DB 4294790 CTATTGCGGACATGAGTATTGACATGCTATGACTGCGG 4294750
RESULT 9
AAS54048/c
ID AAS54048 standard; DNA: 594 BP.
XX
AC AAS54048;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #179.
XX
KW Antisense: ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PE 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
DR P-PSDB: AAU36189.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 7685; 511p; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 594 BP; 100 A; 214 C; 199 G; 81 T; 0 other:
Query Match 32.7%; Score 33; DB 23; Length 594;
Best Local Similarity 63.0%; Pred. No. 3.1;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
OY 15 CGCGCGGAGATATTGGACCGCTGTACCGCGGAGAGAGCGGCGTGCAGAGCA 74
DB 222 CAGCCCGAGTGTGCGCGAGCGCGGAGCGGAGAGGCGTGCAGTGCCTT 163
OY 75 GGGATCGCGCACGTATGA 95
DB 162 GAGATGCGGCTTCTGACGAA 142
RESULT 10
AAT06769
ID AAT06769 standard; DNA: 28598 BP.
XX
AC AAT06769;
XX
DT 15-OCT-1996 (first entry)
XX
DE Sorangium cellulosum soraphen gene cluster.
XX
KW Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;
KW fungicide; pesticide; myxobacterium; ss.
XX
OS Sorangium cellulosum.
XX
PN WO9533818-A2.
XX
PD 14-DEC-1995.
XX
PR 30-MAY-1995; 95WO-IB00414.
XX
PR 08-JUN-1994; 94US-0258261.
XX
PA (CIBA) CIBA GEIGY AG.
XX
PI Beck JY, Gaffney TD, Hammer PE, Hill DS, Lam ST;
PI Ligon J, Ryals JA, Schupp T, Uknes SJ;
XX
DR WPI: 1996-040226/04.
XX
PT New genes for biosynthesis of anti-pathogenic substances - pref.
PT pyrrolnitrin and soraphen, useful for disease control in plants
XX
PS Claim 3; Page 140-158; 190pp; English.
XX
CC This is the soraphen gene cluster from *S. cellulosum*. The gene
CC cluster may be expressed recombinantly to produce soraphen, or
CC expressed in a transgenic plant for disease-resistance.
XX
SO Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other:
Query Match 32.1%; Score 32.4; DB 17; Length 28598;
Best Local Similarity 72.4%; Pred. No. 5.3;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 TCGCCGCGCACCTCGGCGGAGATATTGCGACCGCTGTACGCCGAGAACGGGCG 59
 DB 12650 TCGCAGCGCACCTCGGCGGAGGCTTTGCGACCGCCAGTCCAGGAGAGTGAGCGC 12707

RESULT 11

ID AAT89956 standard; DNA: 28958 BP.

AC AAT89956;

DT 12-MAR-1998 (first entry)

DE Sorangium cellulosum soraphen gene cluster genomic DNA.

KW Soraphen: polyketide synthase; PKS; antipathogenic; antibiotic;

XX transgenic plant; phytopathogen; resistance; ss.

OS Sorangium cellulosum.

PN US5662898-A.

PD 02-SEP-1997.

PF 01-JUN-1995; 95US-0457342.

PR 20-AUG-1990; 90US-0570184.

PR 02-JUL-1992; 92US-0908284.

PR 31-AUG-1992; 92US-0937648.

PR 01-JUL-1993; 93US-0087636.

PR 08-JUN-1994; 94US-0258261.

PA (CIBA) CIBA GEIGY CORP.

PI Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;

PI Ligon JM, Ryals JA, Schnupp T, Uknes SJ;

DR WPI; 1997-447901/41.

XX Protecting plants against pathogens with genetically transformed

PT biological control agent - which expresses all polypeptides)

PT Involved in pyrrolinitrin biosynthetic pathway

XX Ex 16; Column 104-126; 88pp; English.

XX This genomic DNA sequence encodes the Soraphen gene cluster. This

CC sequence encodes two open reading frames (ORF's), the positions of which

CC are not given in the specification. ORF1 is approximately 25.5 kb in

CC size and encodes 5 biosynthetic modules homologous to the erythromycin

CC genes of Saccharopolyspora erythraea. Each module contains a

CC beta-ketoadslysynthase (KS), an acyltransferase (AT), a ketoreductase

CC (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately

CC adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a

CC type I polyketide synthase (PKS) which has antibiotic activity.

CC Transgenic plants containing such antipathogenic genes like those

CC encoded in the Soraphen cluster should have enhanced resistance to

CC attack by phytopathogens.

SO Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

Query Match 32.1%; Score 32.4; DB 18; Length 28958;

Best Local Similarity 72.4%; Pred. No. 5.3; Mismatches 16; Indels 0; Gaps 0;

QY 2 TCGCCGCGCACCTCGGCGGAGATATTGCGACCGCTGTACGCCGAGAACGGGCG 59

DB 12950 TCGCAGCGCACCTCGGCGGAGGCTTTGCGACCGCCAGTCCAGGAGAGTGAGCGC 13007

RESULT 12

ID AAA75299 standard; DNA: 28958 BP.

XX

AC AAA75299;

DT 15-JAN-2001 (first entry)

DE DNA sequence of Sorangium cellulosum soraphen gene region.

KW Pyrrolinitrin; PRNA; PRNB; PRNC; PRND; antibiotic; biocontrol;

KW fungal respiratory electron transport inhibitor; transgenic plant;

KW antipathogenic substance; biosynthetic gene; fungal resistance;

KW phytopathogen; soraphen; ss.

OS Sorangium cellulosum.

PN US6117670-A.

PD 12-SEP-2000.

PF 24-FEB-1998; 98US-0028934.

PR 30-MAY-1995; 95WO-IB00414.

PR 08-JUN-1994; 94US-0258261.

PR 09-OCT-1996; 96US-0729214.

PA (NOV) NOVARTIS FINANCE CORP.

PI Ligon JM, Lam ST, Hammer PE, Van Pee K, Kirner S, Young TR;

PI Hill DS;

DR WPI; 2000-593649/56.

XX Pyrrolinitrin biosynthesis genes and enzymes, useful for producing

PT pyrrolinitrin for increasing plant resistance to phytopathogens, e.g.

PT fungi or nematodes, and for developing microbial hosts useful as

PT biocontrol agents

XX Example 14; Column 121-146; 124pp; English.

XX The present sequence represents a Sorangium cellulosum soraphen gene

CC cluster. The specification describes pyrrolinitrin gene regions. This

CC region encodes enzymes required in the biosynthetic pathway of

CC pyrrolinitrin. The PRNA enzyme catalyses the conversion of tryptophan

CC to 7-chlorotryptophan. The PRNB enzyme catalyses the conversion of

CC catalyses the conversion of monodechloroaminopyrrolinitrin to

CC aminopyrrolinitrin. PRND enzyme catalyses the conversion of

CC and inhibits a broad range of fungi. It is a fungal respiratory electron

CC transport inhibitor. The genes are useful for developing a host as a

CC biocontrol agent, for protecting plants against a pathogen and for

CC producing large, uniform amounts of pyrrolinitrin. The genes are also

CC useful for producing transgenic plants that express antipathogenic

CC substance (APS) biosynthetic genes. The genes and enzymes are useful

CC for increasing the resistance of host plants to disease caused by

CC phytopathogens, e.g. fungi, bacteria or nematodes. The microbial hosts

CC are useful for controlling or inhibiting the growth of a phytopathogen.

SO Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

Query Match 32.1%; Score 32.4; DB 21; Length 28958;

Best Local Similarity 72.4%; Pred. No. 5.3; Mismatches 16; Indels 0; Gaps 0;

QY 2 TCGCCGCGCACCTCGGCGGAGATATTGCGACCGCTGTACGCCGAGAACGGGCG 59

DB 12950 TCGCAGCGCACCTCGGCGGAGGCTTTGCGACCGCCAGTCCAGGAGAGTGAGCGC 13007

RESULT 13

ID AA05287 standard; DNA: 49377 BP.

AC AA05287;

XX

DT 21-MAY-1998 (first entry)
XX The soraphen biosynthesis gene cluster from Sorangium cellulosum.
DE
XX Polyketide synthase; PKS; biosynthesis; soraphen; SorB; SorA; SorB;
XX SorM; biosynthetic module; beta-ketoacylsynthase; acyltransferase;
KW ketoreductase; beta-ketone processing domain; cytosolic agent;
KW antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW biological control; ss.
XX Sorangium cellulosum.
OS
XX
XX Key Location/Qualifiers
FH 383..760
FT CDS /tag= a
FT /product= SorB
FT /note= "gene product highly homologous to the
FT reductase domains of type I PKs such
FT as eryA from Saccharopolyspora erythraea"
FT CDS 927..19874
FT /tag= b
FT /product= SorA
FT /note= "gene product is highly homologous to
FT type I PKs that are known to be involved
FT in the synthesis of polyketide compounds"
FT misc_feature 942..7115
FT /tag= c
FT /note= "module 1 of SorA"
FT misc_feature 7203..12884
FT /tag= d
FT /note= "module 2 of SorA"
FT misc_feature 13455..19616
FT /tag= e
FT /note= "module 3 of SorA"
FT misc_feature 19871..46318
FT /tag= f
FT 19871..46318
FT /tag= g
FT /product= SorB
FT /note= "gene product is highly homologous to
FT type I PKs genes"
FT misc_feature 19870..24556
FT /tag= h
FT /note= "module 1 of SorB"
FT misc_feature 24638..30820
FT /tag= i
FT /note= "module 2 of SorB"
FT misc_feature 30881..35446
FT /tag= j
FT /note= "module 3 of SorB"
FT misc_feature 35528..40114
FT /tag= k
FT /note= "module 4 of SorB"
FT misc_feature 40190..46318
FT /tag= l
FT /note= "module 5 of SorB"
FT CDS 46851..47891
FT /tag= m
FT /product= SorM
FT /note= "gene product is homologous to the
FT methyltransferase from Streptomyces
FT hygroscopicus that is involved in
FT the synthesis of the polyketide rappamycin"
XX
XX US5716849-A.
XX
XX 10-FEB-1998.
XX
XX 14-DEC-1996; 96US-0764233.
XX
XX 24-AUG-1993; 93WO-US07954.
XX 08-JUN-1994; 94US-0258261.
XX 09-OCT-1996; 96US-0729214.
PR

XX (NOVS) NOVARTIS FINANCE CORP.
PA Beck JJ, Hill DS, Liqon JM, Neff S, Ryals JA, Schupp T;
XX WPI; 1998-158369/14.
XX
XX DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.
PT biosynthesis of soraphen useful as antimicrobial agent against
PT phytopathogenic fungi
XX
XX Claim 2; Columns 47-90; 64pp; English.
XX
XX The present sequence contains a cluster of genes that encode polyketide
CC synthases (PKSs) that are involved in the synthesis of soraphen in
CC Sorangium cellulosum. The proteins encoded by the present sequence are
CC SorB, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC which contain a beta-ketoacylsynthase, an acyltransferase, a
CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC processing domains. S. cellulosum soraphen are useful as a cytostatic
CC and antimicrobial agent active against phytopathogenic fungi.
CC Soraphen-producing transgenic plants or biological control agents can
CC also be produced, which may reduce crop losses and nutritional
CC deprivation for local populations in many parts of the world.
XX
XX Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 other:
SQ
Query Match 32.1%; Score 32.4; DB 19; Length 49377;
Best Local Similarity 72.4%; Pred. No. 5.5;
Matches 42; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 2 TGGCCGCGCACCCTGGCGGAGATATTGCGACCGCTGCGCGGAGAGCGGC 59
DB 29253 TGGCAGCGCACCTCGCGCGGAGTCTTGGCCACCGCAGTCCAGGAGTGAGCGC 29310
RESULT 14
AAAT78508
ID AAAT78508 standard; DNA: 44377 BP.
XX
AC AAAT78508;
XX
DT 26-FEB-1998 (first entry)
XX
XX Platenolide synthase gene cluster.
DE Platenolide synthase gene cluster.
XX
KW Platenolide synthase gene cluster; platenolide production; srmg gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS Streptomyces ambofaciens.
XX
XX Key Location/Qualifiers
FH 350..14002
FT CDS /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAM23716"
FT CDS 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAM23717"
FT CDS 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT /note= "ORF3 encodes protein shown in AAM23718"
FT CDS 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAM23719"
FT CDS 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAM23720"
XX
XX EP791656-A2.
XX
XX 27-AUG-1997.
PD

```

XX PF 19-FEB-1997: 97EP-0301066.
XX XX
XX PR 22-FEB-1996: 96US-0012050.
XX XX
XX PA (ELIL ) LILLY & CO ELI.
XX PI Bargett SG, Kuhstoss SA, Rao RN, Richardson MA;
XX PI Rosteck PR;
XX XX
XX DR WPI: 1997-418047/39.
XX DR P-PSDB: AAW23716-W23720.
XX PT DNA encoding Streptomyces ambofaciens platenolide synthase domain -
XX PT for production of spiramycin-related polyketide antibiotics
XX PS
XX PS Claim 9: Pages 8-33; 81pp: English.
XX CC
XX CC This sequence represents the platenolide synthase gene cluster of the
XX CC invention. This sequence is referred to as the smg gene, and was
XX CC isolated from Streptomyces ambofaciens. This sequence encodes the
XX CC multi-functional proteins which direct the synthesis of the polyketide
XX CC platenolide. Platenolide is the basic building block of the macroide
XX CC antibiotic spiramycin. The DNA can be used to produce compounds
XX CC exhibiting antibiotic activity based on the platenolide structure,
XX CC including specifically the macroide antibiotic spiramycin and spiramycin
XX CC analogues and derivatives. Modifications of the platenolide synthase DNA
XX CC sequence can be made so as to change the number and type of carboxylic
XX CC acids incorporated into the growing polyketide chain and to change the
XX CC kind of post-condensation processing that is conducted.
XX SQ
XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
XX
XX Query Match 31.7%; Score 32; DB 18; Length 44377;
XX Best Local Similarity 68.2%; Pred. No. 7;
XX Matches 60; Conservative 0; Mismatches 25; Indels 3; Gaps 1;
XX
XX QY 1 ATGCGCCGCCACCTCGCGCGAGATATTTCGACCGCTGTCGCGGAGAAAGCGGCG 60
XX DB 29404 ATGCGCCGCCACCTCGCGCGAGAGTCTGCGCCACGCG---GAGCCCGCGCAAGTGGGAC 29460
XX QY 61 TGGCTGCGGAGCAGGAGATCGCGCAGC 88
XX DB 29461 GCCCTGCGCGCATGGGCAATCACCAGC 29488
XX
XX RESULT 15
XX AAT80414
XX ID AAT80414 standard; DNA: 44377 BP.
XX AC AAT80414;
XX XX
XX DT 27-FEB-1998 (first entry)
XX DE Platenolide synthase gene cluster.
XX XX
XX KM Tyactone synthase gene cluster; tyIG gene: multifunctional protein;
XX KM platenolide synthase gene cluster; platenolide production; smg gene;
XX KM polyketide; tyactone synthesis; antibiotic; tylosin; ss.
XX OS Streptomyces ambofaciens.
XX XX
XX FH Key Location/Qualifiers
XX FT 350..14002
XX FT /*tag= b
XX FT /transl_except= (pos:350..352, aa:Met)
XX FT /note= "ORF1 encodes protein shown in AAW22606"
XX FT 14046..20036
XX FT /*tag= b
XX FT /note= "ORF2 encodes protein shown in AAW22607"
XX FT 20110..31284
XX FT /*tag= c
XX FT /transl_except= (pos:20111..20113, aa:Met)

```

```

FT FT /note= "ORF3 encodes protein shown in AAW22608"
FT FT 31329..36071
FT FT /*tag= d
FT FT /note= "ORF4 encodes protein shown in AAW22609"
FT FT CDS 36155..41830
FT FT /*tag= e
FT FT /note= "ORF5 encodes protein shown in AAW22610"
XX XX
XX PN EP791655-A2.
XX XX
XX PD 27-AUG-1997.
XX XX
XX PF 19-FEB-1997: 97EP-0301056.
XX PF 22-FEB-1996: 96US-0012078.
XX PR
XX PA (ELIL ) LILLY & CO ELI.
XX PI
XX PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX XX
XX DR WPI: 1997-418046/39.
XX DR P-PSDB: AAW22606-W22610.
XX XX
XX PT DNA encoding Streptomyces fradiae tyactone synthase domain - for
XX PT production of tylosin-related polyketide compounds
XX PS
XX PS Example 2: Pages 110-134; 220pp: English.
XX CC
XX CC This sequence represents the platenolide synthase gene cluster of the
XX CC invention. This sequence is referred to as the smg gene, and was
XX CC isolated from Streptomyces ambofaciens. This sequence encodes the
XX CC multi-functional proteins which direct the synthesis of the polyketide
XX CC platenolide. Platenolide is the basic building block of the macroide
XX CC antibiotic spiramycin. This sequence was used along with the tyIG gene
XX CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tyIG
XX CC gene is the tyactone synthase gene cluster of the invention. The tyIG
XX CC sequence was isolated from Streptomyces fradiae, and encodes
XX CC multifunctional proteins which direct the synthesis of the polyketide
XX CC tyactone. Tyactone is the basic building block of the antibiotic
XX CC tylosin. The hybrid sequence can be used to transform S. ambofaciens
XX CC lacking the smg ORF1 sequence, or S. fradiae lacking the tyIG ORF1
XX CC sequence, so that they can produce polyketides. The DNA sequence can be
XX CC modified so as to alter the type of carboxylic acids incorporated, the
XX CC number of carboxylic acids incorporated and/or the post-condensation
XX CC reactions performed, thereby resulting in novel tylosin-related
XX CC polyketides.
XX SQ
XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
XX
XX Query Match 31.7%; Score 32; DB 18; Length 44377;
XX Best Local Similarity 68.2%; Pred. No. 7;
XX Matches 60; Conservative 0; Mismatches 25; Indels 3; Gaps 1;
XX
XX QY 1 ATGCGCCGCCACCTCGCGCGAGATATTTCGACCGCTGTCGCGGAGAAAGCGGCG 60
XX DB 29404 ATGCGCCGCCACCTCGCGCGAGAGTCTGCGCCACGCG---GAGCCCGCGCAAGTGGGAC 29460
XX QY 61 TGGCTGCGGAGCAGGAGATCGCGCAGC 88
XX DB 29461 GCCCTGCGCGCATGGGCAATCACCAGC 29488
XX
XX Search completed: November 6, 2002, 13:13:21
XX Job time : 1750 secs

```

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

(without alignments)
6625.829 Million cell updates/sec

Sequence: 1 atcgccgcgaactcgcgc.....gcgcacgtgatgactcgcg 101

IDENTITY_NUC
Gapop 10.0 ,

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

```

1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_srs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fm:*
17: em_hum:*
18: em_in:*
19: em_lm:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_pl:*
25: em_pi:*
26: em_ro:*
27: em_srs:*
28: em_un:*
29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgtg_hum:*
40: em_hgtg_mus:*
41: em_hgtg_other:*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AEF217189	AEF217189 Sorangium
2	101	100.0	71989	1	AEF172664	AEF172664 Sequence
3	99.4	98.4	68750	1	AEF10843	AEF10843 Sorangium
4	99.4	98.4	68750	6	ARI930329	ARI930329 Sequence
5	99.4	98.4	68750	6	ARI99551	ARI99551 Sequence
6	99.4	98.4	68750	6	ARI99559	ARI99559 Sequence
7	99.4	98.4	68750	6	ARI99567	ARI99567 Sequence
8	99.4	98.4	68750	6	AR201097	AR201097 Sequence
9	99.4	98.4	68750	6	AR208671	AR208671 Sequence
10	53	52.5	6462	6	AX153794	AX153794 Sequence
11	53	52.5	34071	6	AX153790	AX153790 Sequence
12	53	52.5	42717	6	AX153789	AX153789 Sequence
13	51.4	50.9	6333	6	AX069081	AX069081 Sequence
14	51.4	50.9	9699	1	MSGMYCSYN	MSGMYCSYN
15	51.4	50.9	14326	1	AE007123	AE007123 Mycobacteri
16	51.4	50.9	24292	1	MECY24G1	MECY24G1 Mycobacteri
17	49.8	49.3	344050	1	MEPRPN1	MEPRPN1 Mycobacteri
18	48.8	48.3	442446	1	AE285636	AE285636 Burkholder
19	48.8	48.0	18160	1	AE007186	AE007186 Mycobacte
20	43.4	43.0	30352	1	MYC0409	MYC0409 Mycobacteri
21	43.4	43.0	41171	1	U00010	U00010 Mycobacteri
22	43.4	43.0	332450	1	MEPRPN5	MEPRPN5 Mycobacte
23	43	42.6	6219	1	AEF340167	AEF340167 Streptomy
24	42	41.6	16124	1	AX024384	AX024384 Sequence
25	42	41.6	16124	6	AX024277	AX024277 Sequence
26	39.6	39.2	41321	1	MSGY409	MSGY409 Mycobacte
27	37	36.6	11042	1	MYV045	MYV045 Mycobacte
28	37	36.6	19806	1	AE007024	AE007024 Mycobacte
29	35.4	35.0	14173	1	AE006999	AE006999 Mycobacte
30	35.4	35.0	37840	1	MYV005	MYV005 Mycobacte
31	34.6	34.3	10354	1	AE005083	AE005083 Halobacte
32	34.2	33.9	46166	1	SCF11	SCF11
33	34.2	32.9	49736	1	AEF19998	AEF19998 Streptomy
34	33.2	32.9	110000	2	LMFICHR1B_09	LMFICHR1B_09 Continuation (10 o
35	33	32.7	10477	1	AE012004	AE012004 Xanthomon
36	33	32.7	11930	1	AE004476	AE004476 Pseudomon
37	33	32.7	179837	2	AC127666	AC127666 Rattus no
38	32.8	32.5	14880	1	SC4A9	SC4A9 Streptomy
39	32.6	32.3	107379	1	SHGCP1R	SHGCP1R S. hygroscop
40	32.6	32.3	192960	2	AC098808	AC098808 Papio cyn
41	32.6	32.3	306650	1	SME591783	SME591783 Sinorhizo
42	32.4	32.1	33778	14	AF133953	AF133953 Spodoptera
43	32.4	32.1	14828	1	AE004676	AE004676 Pseudomon
44	32.4	32.1	28858	6	AR004578	AR004578 Sequence
45	32.4	32.1	28958	6	I47766	I47766 Sequence

ALIGNMENTS

RESULT 1					
AF217189					
LOCUS	AF217189	58733 bp	DNA	linear	BCT 09-JUN-2000
DEFINITION	Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.				
ACCESSION	AF217189				
VERSION	AF217189.1	GI:7453554			
KEYWORDS					
SOURCE	Polyangium cellulosum.				
ORGANISM	Polyangium cellulosum.				
	Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangiales; Polyangiaceae; Polyangium.				
REFERENCE	1 (bases 1 to 58733)				

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Jullien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Jullien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Jullien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source location/Qualifiers
1. 58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
<1. 992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VAGHEAFADFGGCVPRVLLYDNLKNAVVERGDAIREFHTLAL
SADYREFPRVAVARGNEKGRVERAIRYREGFEKRAVADGLDLMKROATEMSSAL
DSWEDRARVROAFDDERSVLLRHPTDPFDEHERAEVEKTPARDLMDTSAPL
DTRRRLVVLADLSQVRIADGNOIVATHYRSMRGOQIEOPHELQGVDEKRRAREHR
GIDRLARARSSOAFRLIYAEERGDVNSAIFRLDLDLGAALAEALAEVLENTI
HIGAVQVDRRRERHRLPPPSIYTRGEHALVYTPHSLTYDLAKKDP"
999. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTETKDKLKSGLGFLACMEQLADKFMLEVLAEIERE
RHKRSLERLKNSRYAAEFKPMDFDSNPKCIDREAVDOLYDSYADLFEVTRAYD
AOKPLLTSTNKAFAWDGQVFPAAACVTLVDRLVHRAEYIEAEISYRLKEAKELNAT
RTKORTKKH"
1998. 6263
/gene="epoA"
/gene="epoA"
1998. 6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRIERAEDPIAIVGAGRLPGGVLDLSGFWLLGSRDVT
GOVPAERMDAAAMPDDLDAPGKTPTRTASPLSDVACFPASFPGISPPRALMGSDRAH
LLEVCWEALENAIAPSAVLGTEGVFTIGIPSEYEAALPRTASAEIDAHGDLATM
PSVAGGRISVYLGRPCVAVDTAVSSSLVAVHACQSLRSGECSTALGSLMISP
STYLVNLSKTRALATDGRCAKFAEADGFGFEGCAVYVYLRSLSGAARDRLAVIYR
SAIHDGASSGVLTPNGSSQEIYVLRALADACCAASVGVYEHGFGTLDGPIEIOA
LNAVGLGDRVATPLLIGSVKTNLIGPEVYASGITGLKLVLSLHGOIIPAHHAALN
PRISMGDLRTYTRARTPMPDNTPRRAGVSFSGSGTNAHVLEAPAACTTPAPE
RPAELDVISARTAAADAHARLRDLETPSQCGDVAFSLATTKSAEHRGLAVAT
SSEGLVLAADAAGOTPGVVGIIADSSRGIAFLFTGGAQTGSGKGLDVMPAT
REAFDLCYRLFNOELDRPLREYMAEPAASVDALIDQTAFTQPALFTFEYALAAVPS
MGPELVAGHSIGELVACVAGVFSLEDAVFLVARGLMOLAPGAGMISIAPEA
DYAAVVAHPAAASVSTAAVNGPDVVYIAGAGCPVHAITAAAMARGAATKALHYSHAFS
PLMAPHLEAFGRVAVESYTRRSIYVLSNLSKAGTDEVSTSGCYVRRARREYVRADG
VVALHAAGTFVEVGPSTLGLVPAVLDPARPALLAASSRAGDEPATVLEALGLW

ANGCLYSNAGLEPSSGGRVPLEPTYMGQERWITDKKADDAARGDRAPGAGHDEYK
GAVKGDRRSARLDHPPEPSGRREXVEAAGDFPLEIDEVDRLRLVYTERAPG
LGEVEIADDAAGLSFNDVOLALGWPDLPGBNPLDILGECACAGRIYAVEGVNGLV
VGOPIYLSAGAGFATHVTTSALVYLPQALSSLETAAMPYALITANVYALDGIAPLO
GERVLIHAATGGVGAIVOMAOHVGAEGVHATGATPEKARAYLESIGRVVSDRSRPFV
ADVAPATGEGEVYVNLNSLSCGLIDKSNLRSRGRFPELICKRDCYKANOUGLREPLR
NLSFSLVDLRGMLEPRARVRLFEELGLIAGVFTPIPIITLFIYAVADFRMAQ
AOHLKLVLTLDPEVOIIRPHAGAGSTGDRDLDRLASAAPARAALAEALRTQ
VSOVLRTEPIKGAALFTRLGDSIMAVEENRRIEASIKLSTFTLSTPNNIALLT
ONLIDLATLSLEVENAAENLRAGVQDSFVSSGADQWEIIL"
6260. .10492
/gene="epoB"
6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLNELHEOGVKLAADGERLTOIAPKNAIPLNLIARISEH
KSTILMLRORLPASEIYVAPAREHVPEPLIDGSSYWLGRGATTPVSGIHAVERD
CTDLDYARLSRAFRKYVARHMLRAHTLPDMKQVIEPVYADLIEITDRGDRSREA
RLVSLDAMSHRIYDTERPILYVAVRLEEQTRVLVSLIDINVLDSLSITKDWL
SFYEDPETSLEPVELSRYDYVLALESKRKSEAHQMSDMYKRWKRAVELEPPMLPKAD
PSTLEIRIFRHTQEWLPSDSMSRLKORRKEGRLPTPTGYILAAFSVIGRMSASPEFTL
NITLFRPLVHVRNDITGDFSTVNLIDITDRDSFQORAKRIOEOLAMNDHDS
GTEVORAEARVGIQORGALFPVLTSLNCOGVYTSIORGTPTVYSTORPOLLDH
QIYEHDDGLVLANMDYDGFPPDLDDMLAEVATLRLTERPNEQMRCSIPRAOLE
ARASNETLSLSEHTLGLFARVLEQLPESGAAYVSAKRTLYEELSRLGRKLRL
EGARPNTLVAAMKMGQVVAIVLEQAAVYVADIDAPAEIIVHLHGEVLR
LQPMIDGLKSPMDIOLRLVSDACVEGQDLPMPQIOTSDLVYVYTGSGGLPK
GVMIHGRGVNLTIDINERFEIYVLAIVLSTLSFSDYVDVFIILAAAGIYVPPA
SKIRDPAHMAALIEERKVIYVMSVPALRIVSEHGRPSIARSLSTLSGMDIPV
GLPGELOAIRGVSYISLUGATREASINSGIYVRVNDVSMASITPGRRLNOTPVLD
EALREPVWPGQLYTIGVGLALGTWRDEKRNKSLVHPETGERLYITGDGRLPD
GNIEFMGRDNOIKLGRYKVELGEILETLKSHPNRDAIVYVPGDAANKLLAVVP
EGTRRRAEODASLKTERYDARAHAAKADGSDGRVQFKLARHGLRDLGKPVVDL
TGLVREAGLDVYARRRSYRPLEAPIFPEVERGLSCISVEPGLAPKFRYPSAG
STYPOVTAAYAKSGRIEGVDEGFYHYHPEHRLKIVSHGIERGAHVONDPDEPA.
RGLLVGRIDALIESIXGSLSRFCLLAGYNAQLMEBAPSENTVCYVCGPDEQVYR
PVLDLRHSDVYVHGMGLKVPDRQVQCTLGDGSDSPKATRTGAPGAGDHPADLEI
FLTRKLIDVWPVTEVEDALPLTNGVAVRKALREKRTDPSRSHGTAHPADALEI
IYAAVREYVLEGVLOOSFVLDGATSIHYVMSRLQKRLDREIATIELRQYPMIGS
LASGVRBSKDLQPRNMQDRVEARRKGRRS"
10639. .16137
/gene="epoc"
10639. .16137
/gene="epoc"
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEOESSAIVAVIGSGRPPGARDDEFRNLRDGTAEVQRESEQ
ELAAGVPAVLVDNRYVAGSVLVDVDRPDAPFGISPREALEDMPDHRIEMECWE
ALENAGYPTLAEGSIGYACGAKMSYSITSNLHEPAMRRMPGOWTILGDKOYLAT
HYSTRNLGRPSISYQTKSTSLVAVNHIALKMSLRECDMAAGGIVIRIHRKQYI
AEGGIFSPDGHCRFADANANGTIMNGGVNLIYLRDLDAIDGDEVAVVILSAATND
GARKIGTFAPSRVGOQAIMEALIAVGADEARSIOYIEHNGTTLGDAIETLAALRVD
DDASTRSCAIGSVKTIIGHLESAAIGLAKIVYLALEHQLPSPLESPNSIDF
ASSPYVYTKDMWTGPPRAGVSPFGIGGTNAHVLEAPAKLPAAAPARSAEL
FYVSKAALDAALDAARLRDHOAQGISLGDVAFSLATTSMPHRLAAMPASSEL
REGDLAARGOTPGAVRGCSPGNVPVVFYFQGGQVAMGQDLAEAEVTHAAL
SACDRAIDAEGWSSLABELADESSQIETLDVVOYFVALAVAEALMRWGVGPV
VIGHSMGEVAANVAGALSLEDAVAILICRSRLLRISGCEMAVTELSLEAEARLV
GYEDSVAVSNSPSRIVLSGPAIGVLSLANKGVCFRRVAKVDVSHSPQVPLR
BDLLAALGLRPRRAVPMRSITVYCAMAGPELAINYMMNLROVRAEAYVNOLOG
GHGLFVEKSPHPIITTSYEBKRRAAQNRGAAGVSIKRRQDERPAMLELGLALRGY
VFMGRLPFAGGRVPLPYIPWOREKTYIEAPKASAGDRVRAGGHPLLDEKOTLST

QSTRMWTETLTLRLPWLGDHRYOGAVVPEGAAYLEMAISSGAEALDGIPLQITDVT
LAEALAFAGDAVAVYVYVTEOPSGRLOFOLASRAGPGHSPFHAGALRVRETE
VPAGTILASVABRILAOASMPAAATYAEITMELQYOGAFNOGIAEILMRGEGALRVRETE
DAAGSAAYERLHPALDACCPOYVSLPAGGEGATPWPVPEVGSIRLIORPGEIEMCHA
RVNHGRQTPDROGADENVDSDGAVAAVSGVLAQRI.PGGVNRREEDMDLLEMEP
AAVGTAVNAGRMWLLGGGGGLGALRLMAGAAVAAHAESENSTAGVALLAKEM
DGOAPTVAVHLSIDGGELDPGLGAGALDAPRSADVSPALDPAIVLHRCGCDVLMY
QALAGMRDAPRLMLTLRGAAVAGAVSVYQAPLGLGIVMEHEDLRCARVLD
PTRPDGELGALLAELADDAEVALRGERGECVARIYORPETERPREGIESCVPTDY
IRADSTIYVGTGLGLGLSVAGMLAEKAGCHLYVGSNGASVORAVALEKRGAR
VTAKADVADRQLERLIREVTTSGMPLRGVNAAGLIDDLMOQPPARKVMAP
VQGLAHALREAPLSEFVLYASGVILSGGGNTAAATFIDALAHHRAGGLPA
LSVDMGLFAEYMAAEOEDRGARLSKGRSLTPDEGLSALRLESSEGVYGLPV
PRLMVELYPAASSRMLSLVTAHARASGAGPDGLDLRLRAAEPSAESGLLEPLR
AOISOVLRLPGKTEVDAPLISGMNLSIMGLIELRNRIEAMTIGVPAITLMTYPTVA
LSGHLARACCAAAVESPHTTADSANVELIEMSODLITLAAKRAKT

gene
CDS
/gene="epod"
/protein_id="AA062883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQNPLKQAIIITQRLBERLGLAQLELTERPIATV
GIGRFGGADAPAEFWELLDAERDAVQPLDRNALVGAFAVEAVPHAGLTPEDIC
FDAEFGISPREARSLDPOHRLLEVMAGLEDIPRSIDGSRGTGFEVAFADYD
RTVARLREERDAYSATGNMLSIAGRSLSYLGOGCLTYDTCSSVLVILHACRS
LRAGESPLADAGVSTLSPDMEAAARTQALSPDGRCTEDPAAGSVREGGGIVY
LKRLSDAORODRIMALIRGSAIHNDRSTGLPAPNLAQTYLREAIRSHVNAVAY
DYVTRGTGTSIDGPIEYERLRAVTPARSGTCVGAATNIGHLAAAGVGLIK
AALSLTHERIPRNLEPRLNRIRESALALATEPVMPPTDPRRAGVSGFMSGT
NAHVLEAPAVEILMPAPAPERSAELVLSGSEALDQAARLREHLMHPELGLDY
AFSLATRSANTHRLAAVATREGLIALSVAGGQPPAGAACRIASISRGKLLFT
GOGAOTPGMGKGLCAAMPAPAREARDCTYDRLDRLDPAVMAFASASLIDOT
AFQOPALFAVYATATLALMRSGVPELLVGHISIEYLVAACVAFPSLEDGRVLYAAR
RLMGGLSNGAMWSLGALEAVAAVAAPHAMWISIAVANGEDVYVIAVEAVOALIA
GFAARGVTRKLIHVAHSPLMPEMLEEPRVAASYTRRPSVLSVNSLKGKVTDE

Query Match
Best Local Similarity 100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCGCCCGCCACCTCGCGCGAGATATTGGACCGCTGTGACCGCGAGAGCGGGCG 60
|||||
Db 36150 ATCGCCCGCCACCTCGCGCGAGATATTGGACCGCTGTGACCGCGAGAGCGGGCG 36209
|||||

Qy 61 TGGCTGCGCGAGCAGGCGATCGCATGACTCGCG 101
|||||
Db 36210 TGGCTGCGCGAGCAGGCGATCGCATGACTCGCG 36250
|||||

RESULT 2
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothonones C and D
JOURNALS Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
Location/Qualifiers
1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 101; DB 6; Length 71989;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCGCCCGCCACCTCGCGCGAGATATTGGACCGCTGTGACCGCGAGAGCGGGCG 60
|||||
Db 36150 ATCGCCCGCCACCTCGCGCGAGATATTGGACCGCTGTGACCGCGAGAGCGGGCG 36209
|||||

Qy 61 TGGCTGCGCGAGCAGGCGATCGCATGACTCGCG 101
|||||
Db 36210 TGGCTGCGCGAGCAGGCGATCGCATGACTCGCG 36250
|||||

RESULT 3
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothonone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangineae; Polyangium.
REFERENCE
1 (bases 1 to 68750)
Mojnar,I., Schupp,T., Ono,M., Zitzke,R., Mlhamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Stratanu,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A.J., Goff,S., Schmid,J. and Ligon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothonones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE
20130945
PUBMED
10662695
REFERENCE
2 (bases 1 to 68750)
Mojnar,I.
Direct Submission
Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultures Research Institute, Inc., 3054 Cornwalis Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AA026904.1"
/db_xref="GI:6724238"
/translation="ASIDALFARATSAVRLDDHGRAPRHHVLAENGIEDLRALREH
LRIDGSPSEFCMLGIDTVYELAHNDPLASTSHHARSLRHPWTDAMVOPRAY
RWLAARGAPFLREYEBERARTAOERRLWLAAPCPAPDLPREDNDNGLPLEP
MSPEVAERLRASVATPELCAALAWLTGAGPSGAPAYMLEPILNLLGFLPT
AIAAASAPGTSEALRGAAARLFASMEVAVSSKSGQGNIPALMERLRTIYAMGADN
LSRPERAIAAEVRLRAOPAPPAAGGLAVAGVSSGRLSGVTDGALYSGDND
IVMFOGRISPVVLVAGDPEPELAPLSQMLFAHNAGTISKVLTEGSLIYMANR
QARPMSLYHARGFMAVYQNAVPPDEGAPRYVORSTIMEPHEPTPCILPAGSARS
LACDEHLYWCELSAGRLERHHPHHRGASRAYIGEPITATYPSLTLNATHYL
WADPDRAIILGVDRGTVEPIVLAETHHPPAHVSESDNDIFALTGDPDSRWHEHIL
SGASTVADYORQLMDRDPDWVLRNRCLEFTTNDRIILARS"
complement(1900..3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to DD-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA026905.1"

/db_xref="GI:6724239"
/translation="MGALISVAAPGALGAEDEEGOPGODAGAGALAPAREVMAAEVA
ACQMPAIVLVAAGDDVDVADAVTELGSAPMRDITIFRIASMTKAVATAVMLVE
EGKLDLSDSPVDRMLPELAPNRKVLAIDGPIDETVPERPITVLDLMTFMGEGISDA
SSPIORAIIDELIGLVNAOPVPMTPHGPDEMRIRICGTPLMHQPGACOMYNTGSLVQCUL
VGAADQGFDAVYRERILAPLGRDITDFHYPADKLARFAGCCFTTDEQTEGKTRMD
GAEASVAPSPAPFSGAAGLVSTVDVYLLFARMMLNGVHGRLLSAASREMTACHL
TPAQKASSFFPFPFETHGCMGMAVVTAPDAVSEVPGRGYMGDGGFTSMINDPGREL
IGVMTOSAGFLFSGALERFMSVYATESA"
3415..5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGTEROVLISVLTALITVYARASGEIARLROPEVIGELFG
GVYLGPSVVGALPGCFHRALEPFAVGVLSIGTIGALLILMAGIEVDGILRREA
RPGALSALGAIAPLPAAGAFSALVLDRLPSGLFGIVLSTVAVSVIAKVLTERSM
RSYAGVTLAAGVSEVAAVLAAMTSSSGASPALAVARSAALSGFLFMVLVGR
LTHLARMVADATRVSGOVSLVLTFLAALTORLGHPLGAFALGVLNVASPR
NRLDGVOTIVAGLEAPVFEVLGMRVDSOLRTPAANGTVALLATATPAKVVYAA
LGARLGLGSGSAALVAGLNKMGCTDLVATVYVGLGILSNATMYAVVALVYTA
SPALILTEKRAPITOEESARLEKEBAARATIPGERILVPIVAHALGFAFDIVES
IVASKRKILEYVDITELSVYEQAPGPSRAGEASNLARLARGILRVIRORRELGS
IOAILASRDHDLVIGARSPARAGMSFGLQDAIVORAESNVLVVGDPPAERAS
ARRILVPIIGLEYSFAADLAHVALAMDAEVLSSAOTDPCGAVVMRDRERSRAY
ARSVEAVEPGRGLGVRVSRVYHGVAHPSDETRRLARAPVDLVLGCVDHGPIGLRL
YLGSTVESVYVSRVVALVHAGTREQVR"
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGTRCAIAEYIPEGSGNPEPQMTSHECTILNASDRDA
QVATVYFSDRDPAGPYRVTVPARKTRHVFNDLTPEPIPRDTYASVIESDAPIV
OATRLDSROEADMLSTIAYTDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVDTGRRGRIRIGHTIGLASMALAGCGGSEKTVGTRAP
GADARTADVDDAATTRLAVDVYHLSPPERLEAGSERVYVWRPSPERMVAVLD
YNADSRGKLAETVYVYANFELLITAEKSSPQSPSAVIGTISVG"
7610..11875
/gene="epoa"
7610..11875
/gene="epoa"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRIERAEDPIAIYASCSLPGCVIDLSGFTLLGSRDY
GVPAERMDAAMFDDPDAPGKTPTRASFUSDVACFPDASFEGISPRALRMDPNR
LLEVCWEALENNALIASALVGTETGVFIGIPSEAEALPQATASAEIDAGLGTM
PSVAGRIYALRGPCYAVADTAVSSSLVAHLACQSLRSGECSTALAGSLIMP
STLVMLSKTRALARDRGKAFSAEADGFGREGCAVYVLRKLSGARADDRLIAYVQ
SAIINHGAASSGLVPMGSSOETVLRALADAGCAASGVYVAHGCTGDIETIQA
LNAVYGLGDVATPLLIGSKTNLGHPEASGITGLKAVYLSIQHGQITAHILHAOLN
PRTISGDLRLVYTRATPWPDMNTPRRAGVSSFGMSGTNAHVLEAPATCTPPAE
RAEELVLSARTASALDAQARLRDHLLETYPSCUGDAVAFSLATRSAMEHRLAVAT
SREGLLAALDAAGCOTSPGAVRSIADSSRGKLALETGQAGOTLGMGGLVDVMSAF
REAFDLVRLFNOELDRPLREVMMAEPASVDALLDQTAFTQPALPTEFYALAAIIRS
KGVEPELVAGHSIGELVYACVAGVSTEDAVFLVARGIKMALPAGCAMVSTAEPA
DVAAAVAPHAASVSTAAVAPQOVYIAGAGQFVHAITAAAMARGARTKALHSHASH
PLMAPLEAFGRVAESVYRSPSIVLNSLGSACTDEVSPPGVYRRAHREYVRADG

VKALHAAGATEVENGPKSTLLGLVPACMPDARPALLASSRAGDEPATVLEALGLM
AVGVLGSMAGLPPSGRRVPLPTTYPMQREKRYIDRKADDAAGDRAGAEVEEG
GAVRGDSARLADHPPESGREKREKVAEDFRLEIDEVPLHLVLRVYERAPG
LGEVEIYAVNAGLSENDVOLTAGLPPDRLPGKNDPILLGSCACARIYVAGGVGLV
VCOPIYALISAGAFATHTVTSALVLPQRQALSAITAAAMPVAYLTAVALDIAALQ
GERVLIHAATGGVGLAIVOMACHVAGAEVATGTEKRYTESLVRVYSDRSRFV
AVURAMTGGEEVDVYVNLISGRLIKSEFNLISHRFPFELGRDCYANOGLRFLR
NLFSFLVDLRGMLEPARVRLLELLEGLLIIAGVTFPPPIATLPIARVADAFRMAQ
AOHLGKVLVLTLDPEVOIRIPTHAGAGPSTGDRDLDRLASAPARAALAEAFRTQ
VSQVLRPEITKGALEAFTRLGMDISIMAYELRNRIEASIKIKLSTTFTLSTSPNMLLA
QNLDLALATALSLEKVAENLNAQVONDFVSSGADQDEITNL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P: contains condensation/meterecyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTNOLNLELHOGIKLADGERLQIQAPKVALNPMLLARISEH
KSTITLMLNORLPASIVAPAPERHAPPLDIOESYMLGRGATVPSGIHAYREYD
CTDLVPRLSRAFRKRVARHDMLRHNTLPDMQVTEPYKDADEITIDLGLDRSTREA
RLVLRSDAMSHRILYDTERPLVHVAVARLDEKOTRLVSLIDLVNGLSGLTFKMDL
SFYEDPETSPLVLELSYRDYVALLESRSKSEAHOSMDYKRIAEPLPPRLPKAD
PSTLKEIRFRHTEQMLPSDSGRKLRRRGECRLTPTCYILAFSEYTRKWSAPFTL
NITLFNRLEVHPRVNDITGDPVTSWVLDDITDRDSFQQRARIEQMLEWADHCDVS
GIEVOREAARVLIQRGALFPPVYLSALNOQVGVTSQRLGTPYRSTQRPQLDLH
QLYEHNDGDLVAMDIVDGVFPDLDDMLAEAVVYLRRLTEPMGEQVCSLPPAOLE
ARASANTNALLSEHTLHGLFAAREQOLPMQALVYASAKTILYEELSSRSRGLARL
EGCAPRNTIYAVMEKMGEOYAVIYAESGAAYVPIDADLPAERIHVLDHGEVKV
LTOPMLDGLKSNPFGIQRLLVSEAGVEBGDQPPMPPIOTPSDLAYVYITSGSTGLPK
GVNIDHRGAVNTILIDINERFETGPDDRVALTSLSFDSLVDVFGILAAAGTIVVDA
SKLRDPAHMAELIEREKYVVMNSVPALKMLVEHREGSPDSLARSRLSLSGDMPV
GLPGELOAIRPGVSVISLGATASISIGYVRVAVDLSMASIPYGRPLRQOTFVLD
EALDEPRVWVPGOLYIGVGLALGWBRDEKRSSELVHPETGERLTYKTGDSGRYLPD
GNIEFMGREDNOIKRGYRVLEGETEETLKSHPNVYDAVYVYVGNDAANKILLAAVVP
EGTRRRAEQDASLATERIDARAHAAEADGSLDGRVOPFKLARHGLRDLDBKPYVDL
TGQDPREAGLDVYARRRSYRTLEAPITPVERGRLLSCLISVERPDAVLKPRFPSAG
STYPVQTYAAVYKSGRIEGVDEGFYVYHPEGRLLLSHGERGHAVNQNDVDEEA
FNLIVGRIDALIESLJGSSSREFCLLEGYMLQMEAPQNLINVCVQGFNEQVOR
PVLDLRSHDVYVYHGLMGGRVDPROVOCTLGDSSSPRATRTGAPRGEOHFNEDLR
LVRTKLPEYMPVTFVELDAPLITNSGVDRKALRERDTSPPRSRGHTAPRDALEET
LVAVYREVYGLSEVGLQSSFYVDLQATSHIYMKSLKRLDRLEIATITELFOYPLGS
LASGLRDSRDLDOHRNMODRVEVRKGRRS"
16251..21749
/gene="epob"
16251..21749
/gene="epob"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match

98.4%; Score 99.4; DB 1; Length 68750;

Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 41862
|||||

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 41862
|||||

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

Db 41762 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 41862
|||||

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 41862
|||||

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 60
|||||
Db 41762 ATGCGCCCGCACCTCGGCGGAGATATTGGACCGCTGTGTACACCGAGAACGGGGCG 41821
|||||

QY 61 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGACAGGAGATCGGCACGTGATGACTCGCG 41862
|||||

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn., and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No.3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ATGCCCGCCACCTCGCGCGGAGATATTGCGACCGCTGTACGCCGGAAGCGGCG 60
|||||
Db 41762 ATGCCCGCCACCTCGCGCGGAGATATTGCGACCGCTGTACGCCGGAAGCGGCG 41821
|||||
OY 61 TGGCTGCGGAGCAGGAGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGCAGGAGATCGCGACGTGATGACTCGCG 41862
|||||
RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn., and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No.3.4e-14;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ATGCCCGCCACCTCGCGCGGAGATATTGCGACCGCTGTACGCCGGAAGCGGCG 60
|||||
Db 41762 ATGCCCGCCACCTCGCGCGGAGATATTGCGACCGCTGTACGCCGGAAGCGGCG 41821
|||||
OY 61 TGGCTGCGGAGCAGGAGATCGCGACGTGATGACTCGCG 101
|||||
Db 41822 TGGCTGCGGAGCAGGAGATCGCGACGTGATGACTCGCG 41862
|||||
RESULT 10
LOCUS AX153794 6462 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 118 from Patent WO0140497.
ACCESSION AX153794
VERSION AX153794.1 GI:14535430
KEYWORDS
SOURCE unidentified bacterium.
ORGANISM unidentified bacterium.
REFERENCE 1 (bases 1 to 6462)
AUTHORS Jeanin,P., Pernodet,J.L., Guerinneau,M., Simonet,P., Courtols,S., Cappellano,C., Francou,F., Raynal,A., Ball,M., Sezonov,G., Tuphile,K. and Frostegard,A.
TITLE Method for obtaining nucleic acids from an environment sample, resulting nucleic acids and use in synthesis of novel compounds
JOURNAL Patent: WO 0140497-A 118 07-JUN-2001;
Aventis Pharma S.A. (FR)
FEATURES
source Location/Qualifiers
1..6462
/organism="unidentified bacterium"
/db_xref="taxon:2338"
BASE COUNT 1073 a 2067 c 2135 g 1187 t
ORIGIN
Query Match 52.5%; Score 53; DB 6; Length 6462;
Best Local Similarity 70.3%; Pred. No. 0.0042;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
OY 1 ATGCCCGCCACCTCGCGCGGAGATATTGCGACCGCTGTACGCCGGAAGCGGCG 60
|||||
Db 4750 ATGCCACAGCGTGGCGCGGAGATCTTGGCTACTGCCGGAGTCCGGAACACAGCG 4809
|||||
OY 61 TGGCTGCGGAGCAGGAGATCGCGACGTGATGACTCGCG 101
|||||
Db 4810 TATCTGCGCTGCTGGGATCGGCATGCTTTCGATTTCGG 4850
|||||
RESULT 11
LOCUS AX153790 34071 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 114 from Patent WO0140497.
ACCESSION AX153790
VERSION AX153790.1 GI:14535426
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 34071)
AUTHORS Jeanin,P., Pernodet,J.L., Guerinneau,M., Simonet,P., Courtols,S., Cappellano,C., Francou,F., Raynal,A., Ball,M., Sezonov,G., Tuphile,K. and Frostegard,A.
TITLE Method for obtaining nucleic acids from an environment sample, resulting nucleic acids and use in synthesis of novel compounds
JOURNAL Patent: WO 0140497-A 114 07-JUN-2001;
Aventis Pharma S.A. (FR)
FEATURES
source Location/Qualifiers
1..34071
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="insert d'ADN du cosmid a26c1 - brin codant"
BASE COUNT 5791 a 10858 c 11089 g 6333 t
ORIGIN
Query Match 52.5%; Score 53; DB 6; Length 34071;
Best Local Similarity 70.3%; Pred. No. 0.0036;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
OY 1 ATGCCCGCCACCTCGCGCGGAGATATTGCGACCGCTGTACGCCGGAAGCGGCG 60
|||||
Db 22973 ATGCCACAGCGTGGCGCGGAGATCTTGGCTACTGCCGGAGTCCGGAACACAGCG 23032
|||||
OY 61 TGGCTGCGGAGCAGGAGATCGCGACGTGATGACTCGCG 101
|||||
Db 23033 TATCTGCGCTGCTGGGATCGGCATGCTTTCGATTTCGG 23073
|||||

RESULT 12
LOCUS AX153789 42717 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 113 from Patent WO0140497.
ACCESSION AX153789
VERSION AX153789.1 GI:14535425
KEYWORDS
SOURCE
ORGANISM synthetic construct.
artificial sequences.
REFERENCE
AUTHORS Jeanin,P., Perinodet,J.L., Guerinneau,M., Simonet,P., Courtois,S.,
Cappelletto,C., Francon,F., Raynal,A., Ball,M., Seconov,G.,
Tuphile,K. and Frostegard,A.
TITLE Method for obtaining nucleic acids from an environment sample,
resulting nucleic acids and use in synthesis of novel compounds
JOURNAL Patent: WO 0140497-A 113 07-JUN-2001;
Aventis Pharma S.A. (FR)
FEATURES
SOURCE 1..42717
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="cosmid a2691 brin non codant"
BASE COUNT 8230 a 13520 c 13184 g 7782 t 1 others
ORIGIN
Query Match 52.5%; Score 53; DB 6; Length 42717;
Best Local Similarity 70.3%; Pred. No. 0.0035;
Matches 71; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
Qy 1 ATCGCCGCCACCTCGCGCGAGATATTGGACCGCTGTACCGCGAGAAGCGGCG 60
Db 11147 ATCGCACAGCGGCGCGCGCGAGATCTTCGCTACTGCCGAGATCCGAAAAACGAGCG 11088
Qy 61 TGGCTGCGGAGCAGGAGGATCGCGACGATGATGACGCG 101
Db 11087 TATCTGCGCTCGCTGCGCATCGCGATGTTGCGATTCGCG 11047
RESULT 13
LOCUS AX069081 6333 bp DNA linear PAT 25-JAN-2001
DEFINITION Sequence 61 from Patent WO0102555.
ACCESSION AX069081
VERSION AX069081.1 GI:12578964
KEYWORDS
SOURCE
ORGANISM Mycobacterium tuberculosis.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
AUTHORS Glouel,B., Guilhot,C. and Camacho,L.
TITLE Method of making and identifying attenuated microorganisms,
compositions utilizing the sequences responsible for attenuation,
and preparations containing attenuated microorganisms
JOURNAL Patent: WO 0102555-A 61 11-JAN-2001;
INSTITUT PASTEUR (FR)
FEATURES
SOURCE 1..6333
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
/note="mas gene, mycocerosic acid synthase."
BASE COUNT 1010 a 2010 c 2202 g 1111 t
ORIGIN
Query Match 50.9%; Score 51.4; DB 6; Length 6333;
Best Local Similarity 69.3%; Pred. No. 0.01; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
Qy 1 ATCGCCGCCACCTCGCGCGAGATATTGGACCGCTGTACCGCGAGAAGCGGCG 60
|| ||||| | || ||||| || ||||| || | || ||||| ||

Db 4732 ATTCGCCGCCACGAGGAGATTTTCGCGACCGCGGCAATCCCGAGAGAGCC 4791
Qy 61 TGGCTGCGGAGCAGGAGTCCGCGACGTGATGCGCG 101
Db 4792 ATGCTGCGGACATGGCGTGCAGCATGTCTACGATTCGCG 4832
RESULT 14
LOCUS MSGMYACSYN 9699 bp DNA linear BCT 26-APR-1993
DEFINITION Mycobacterium bovis mycocerosic acid synthase gene, complete cds.
ACCESSION M95808
VERSION M95808.1 GI:149977
KEYWORDS mycocerosic acid synthase.
SOURCE Mycobacterium tuberculosis (strain BCG, pathovar bovis) DNA.
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
AUTHORS Mathur,M. and Kolatukudy,P.E.
TITLE Molecular cloning and sequencing of the gene for mycocerosic acid
synthase, a novel fatty acid elongating multifunctional enzyme,
from Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin
JOURNAL J. Biol. Chem. 267 (27), 19388-19395 (1992)
MEDLINE 92406887
PUBMED 1527058
FEATURES
SOURCE
Location/Qualifiers
1..9699
/organism="Mycobacterium tuberculosis"
/strain="BCG"
/db_xref="taxon:1773"
complement(153..1049)
/note="ORF"
/codon_start=1
/transl_table=1
/protein_id="AA25367.1"
/db_xref="GI:149978"
/translation="MDYEQMDGVAITITMSQLRRLTNVARELSRGSSTGDRVYSK
PQGLEVVAFGLAQAGRIAYPLSPGGVTVDESDSVLSSEPAIITTSADVAV
QNVARRPGEPSPIITENDLDLDAPNGYFEDYEPATYQYISGSTRTPAGVMSH
ONVNRVEQLMSGYFADTDGIPPNALVSWLPYHDMGLVIGICAPVGLGYPAVLTS
PVSFLORPARMHLMASDFEAFSAPNFAELARRTTDDDMAGDGNILITLSGE
RVOAATIKPFDALLASICRGG"
738..1556
/note="ORF"
/codon_start=1
/transl_table=1
/protein_id="AA25368.1"
/db_xref="GI:149979"
/translation="MLHDVYHGRCSNGHRRRTTORIGTFITGNALMNRKRGDAA
GLQRAEKGDVLESILRFRDHAIVRGTTAQLPCHIQSPVQLNPRROGYRAVPLVY
IHKRGVRLMQLQTRIRMAAGRERTLMKITLPALGVAVAVGCSRSAVLAKSHLS
SAGVCRFLMGPSPDVGCDVAPPRGRRLTQLRICARAGTENNCGATDVYLRGRSP
MTAPNWLALPVEVETIRMSNPATSPRLPQLHAAPMWHLPSEVYRLGRHMGSC"
1741..8073
/function="fatty acid elongating enzyme"
/evidence="experimental"
/transl_table=1
/product="mycocerosic acid synthase"
/protein_id="AA25369.1"
/db_xref="GI:149980"
/translation="MESKVTAVAVIGMKRLPGINSPPDKIMESLREKDDLVTEIPDP
RMDADDDYDPEPGVGRSVSRWGFLEDDVAGFDEFGISEREAITSIDPQRLLETSS
WEALEHAGIDPASLAGSSTAVFTGLTHEEDYLVLTAGGLASPVYVGLNNSVSGRI
AHTGLGHPATMPTPACSSGLMAVHIAFLACRDEBDALAGCAVLLEPACVAAQA
OGMLSTGRCHSPFADADGPFVRSFGCAVILKLRLPDALRDNRIIPAVYRGRTANODR
TEITIMPEDAQVAYYRAALAAAGVOPETVVYVAHAGTPIGTPIETRSIARVYAGC
TPCALGSAKSNMGHSTASAGTVGLIKALITSRHGVAPPEPLFLNFNLPDLSEVGLFY
PQAVTPWPNMGNDHPKRYAVSSFGMSGNVNAIYEEAPAEASAPESGDAVEPRLF
MLSTSSDALROTAROLATVWEHODCVASDLAVYTLARGRRHPVRYAVAAVNPDL
VEGLREVAADGALVDAVGHDRGPVWVFSGGSGQMAMGTOLLASERVPATIAKLE

PIAAESESFTEAITAAQOTVGIDKVOYAFQVALLAAMEOTYGVRGAVGSHM
GSAAYAVGALSIEDARVICTCRSKITMTRPACGANGSVETPAKQONSELMARGID
VYVSVAVSPOSTVIGTSIDVIRLLIARMEQDVAKAREAVDVASHSQVOD IIDDLLAA
ALADIAEMTTPVYVYTSATLFDPRBQPCDGVAVWMDNIRNTYQFAAQQAMGDAVYF
ADVSHPLTLTHAVEQOTGSLDMSVAALANGREOPLEPLHGLTLEHRAAGALDYSA
LYPGRLVADPLPAMTHARLFIIDDGQOORAGACTIVPHLIGLSHVRLIEEPRHW
OGDQTSMLSDHQHYNVALPGAAICEMALAAEVEGEALAEVDTPEOMLLD
BOTPIDAVALSIDAEVNVETVETNRDGETTRHATVALALAAEDOCPPGYDITALLQAH
PHAVNGTMRSSFEARGYVTLGAAFGVLTATTPREPRRCNRCRCRSGSSNAPE
STRMCTVLSRSRASSRYGMMPPAAVAGQOACPLRAYQOCPLLHAAQDQJORD
PRMEDLVIDHEGTVLLAVRGLMGTSQDSRDRSLSSRLTGLMOORALPEVDD
GEASWILLIDPNSAVDPDMALSTLTALAKSHGQVRETCBALSISYDTPENDACGLE
KLQSOLRCDQVVIYVGPBGDPDEHSLLAGREOVRLHVTRTRELAEFEELPPLFVY
TRQADIVYPHDSGERANLEOAGRLVLTISEHMPMLTIDVETDIVERAAOOLL
SGSEDETEAMRGDMYVARLPSPLEGEHERTVALPDHDHMOVQRPDQITLEEV
ASDRPPEGQIEVAWSMSTINFADVILAEGRVETIIDDREOQLGMDPVGYTANGEV
TGHQVGDVGVGFSBEGCCHRTFTCDANLANATVLPPLGLTDEQATPLANTAHATANGV
AOIAGQDVLLSHATGVGQAALISIAARKKAEIATGATGNPAKRAMLDKMCVEHYHYSR
SVEFEELTRRQTDGVDIVLANSRELLAARVLELFPFGGEFVEIGIADVYGRILGL
PFRFGLFEYFLLDLALMSVTPQDPRNELLAVFLLTADGVYTAQCHYPLAAEADAI
RAMSAAHETGTVLIDVPPRSGRSAVNPDEQRLYRBOGSYITIGGGLGCLFSPASKIA
AAGCRVILYTRSQPNPKAROTIEGLRAAGADIVEEGNCIAEPTOLATVLSAATATGL
PLRGVLSHAAVEADATLTINTELIDIRMSQKYGVSNNLHATAGPLDMFCLEFSSCA
ALUGSPGQASAVANSWVDYFAHMPRAQGLVYSIANGCAMEVEYRAFLFEGGINTT
PEEGVGAETVLRHDRAVYSGYIPLIGAMPVLADIVSRBPMGEMFASQTOGRSGKPEM
ELLSLPOBEMAGRRLRLLEBASVILKRTIDADNSFIETGLDLSIGHEMKTTHYETEG
IRLPLKVIATIKTIRALAQYITLADLAEQAAAAPAS"
complement(8468..9532)

BASE COUNT	1691 a	2977 c	3280 g	1751 t
ORIGIN				

Query Match	50.98;	Score 51.4;	DB 1;	Length 9699;
Best Local Similarity	69.38;	Pred. No. 0.0097;		
Matches 70; Conservative	0;	Mismatches 31;	Indels 0;	Gaps 0;

QY	1	ATCCCGCCGACACCTCGCGCGCGGAGATATTTGGCACCCTGCTGCGCGCGGAGGAGCGGCG	60
Db	14649	ATTGCGCGCGCGCAAGGAGGCGGAGATTTTCGACCCCGCGGCAATCCCGCGAAGGAGGCGC	65229
QY	61	TGGCTGGCGCGACAGGAGGATTCGCGCGACGTCATGACTCGCG	101
Db	65229	ATGCTGCGCGACATGGCGCTTCGAG6CAATGCTTCACGATTTGCGG	6569

AE007123/c	AE007123	14326 bp	DNA	linear	BCT 27-APR-2001
LOCUS					
DEFINITION	Mycobacterium tuberculosis CDC1551,	section 209	of	280	of the
	complete genome.				
ACCESSION	AE007123	AE000516			
VERSION	AE007123.1	GI:13882792			
KEYWORDS					
SOURCE					
ORGANISM	Mycobacterium tuberculosis CDC1551.				
	Mycobacterium tuberculosis CDC1551				
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
	Actinomycetales; Corynebacterineae; Mycobacteriaceae;				
	Mycobacterium; Mycobacterium tuberculosis complex.				
	1 (bases 1 to 14326)				
REFERENCE	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,				
	Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Hatt, D., Hickey, E.,				
AUTHORS	Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,				

TITLE	Whole genome comparison of <i>Mycobacterium tuberculosis</i> clinical and laboratory strains
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 14326)
AUTHORS	Flaischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umeyang, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterpack, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Direct Submission
JOURNAL	Submitted (23-Apr-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. 14326

/organism="Mycobacterium tuberculosis CDC1551"
 /strain="CDC1551"
 /db_xref="taxon:83331"
 /note="clinical strain"
 complement(112..6447)
 /gene="MT3010"
 complement(112..6447)
 /gene="MT3010s"
 /note="similar to GB:U00010 PID:466793; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="mycocerosic acid synthase"
 /protein_id="AAK47337.1"
 /db_xref="GI:13882793"
 /translation="MESRYTPVAVITGCGCLPGGINSPPDKLMSLLRGDLVTV
 RMDADVDPEPGVGRSVSRWGGFLDVGAFPEFGISERATSIDQQR
 MEALHEALDPLASLGSSTAVFTGLTHEDLVLTTRAGLASPVATGLNNS
 AHTEHLHPMTFPLACSSGSLMAHLACRSLHDGEADLALAGCAVLEPRHAS
 OGMSLSTRCSPADADGFEVRSBGCAEMVTLKRLPDLRNGNILEFVSTIAR
 TETLMPEMDQAVAYRRAALAAAGPEPTVYEAAGTGPIDPIEYVSLAR
 TPCALGASKMKMGSTASAGTYLKAITLNRHGVPLHEPRLLDELSVE
 POAVTPRNNGDHTPKRYAVSSFGMSHTLILYERPAEASPESSPGTAA
 MLSTSSALCKDTRQLATVWEEHODCVASDLAVTLERARRPRTAVVAA
 VEGLEADGGLADVAAGHDGRPWVEFGQGSOMAMGTOLASEVPEFAN
 VPIAESEFSTVEAITAQOFTVGTIDKVPQVAVALATMQEOTGVPEFAN
 GESAAVAAGSLSEDAARVLCRSKLMTRLAGAGAVSELPKQVNSELMMA
 VYSSVAASGALSTVIGGTSDFYRDLIARWQRDMAGAREVAVDASHPOYDIL
 ALADAPMTPRVPTYSATIDPPREPQCDAAWYVDNRNTPQRAAVQAMAE
 AELSHPLTLTAVEOTGRSLDMSVAALAGRRREPDLHRLGLTSLRAGAA
 LYPARLDALPAMTHARLFLEIDDDGOEORAGACTIVPHLIGSHVRLTEEB
 OGDVTSVLSMDIOVHNVALEPGAAYCEMALAAAEVGEAAEVRDITFRO
 EOTPIDAAVSIADGVNFTVETNRDETRRHATAIRAEDDCPPGDIITPA
 PHAVGTMRSPFAERGVTLCGAAGLTTAHTBAGAAVLAVALPASTIRCP
 IHPALLDLCPOGACVAGTAGGILLPTGVSRILAVGPTRARACVRLRK
 TRGCAADVDYDERGYTLAVRGLRMKGTGSENDEDRRLYSEMLITLMQQR
 DEGAOSWLLDITSAVDTPMDLASTLDLAKSHGPGTEBASLSMSVQDTPP
 BKLGSVGRGCVIYVPRVGPDEHSLAGSHOYNHVLRITRELAEEGELE
 VTRQOIKPMDSEERANLEQAGRLGLRVTISEPHRLTTLIDVDEHNDVER
 LSGSEEDTPHRSNDMWYARLTPRLPHEHRRATVLDPPHGDGVORRPPG
 VASDVPVPGCOILEVAMSSINADVLIANGRPVLTIDREPDGMDGVYVTA
 VTGRGVGVNCGFESEGGCMRTFLITCDANLAVTPRGITDEQATTAATATA
 LAQIAGKGVLIHSATGCGGAAISINAKGAELFNAGAPARALAKMGVAV
 RSVSEADLEIRDTDGVDIVLNSITGELAAAGLELAFEGREVEIGKADYV
 LEFERGLTEFYEDLALMSVYQDPRYVELATFEKLLADVLAPACTYRPLA
 TRAMNAHTEKLVLDVPRSGRNVATPCQALPYRDGSGYITGGIAGLGEV
 AAAGGRVLTARSOPDKARORTTEGIRAGAADIVCGNIAIDPPIADVLSA
 LPLRGVLSAAVVEDATLNTIDTELIDRMSPRVRSNMRHRTLGQPLDNIC
 AALGSPOQGVAAANSMVDFAHMRAGQCLPVSAIANGMEGVAGTLAGGSC
 TPEBAVAFETLVHDAVYSGYITLCAAPQDLVRSRPMGEAFSTGORSRGS
 MELLISPDDEWAGRLRLLEVOASVLTIRRTIDDRSTIEGDSIGLEMTHTH
 GILRLPKVIAIINNTARALAOVLADTIAEEQAAAPAS"
 7067..8809
 /gene="MT3011"
 7067..8809
 /gene="WT3011"
 /note="similar to PID:1658531; identified by sequence

```
/gene="MT3011"  
/note="similar to PID:1658531; identified by sequence
```

```
similarity: putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA synthase"
/protein_id="AAK4738.1"
/db_xref="GI:13882794"
/translation="MSVRSLLPALRACARLOPHDPATFMDYEDQMDGVAITLWMSOL
YRRLTNAQELSCSGTGDRVVISAPQGLEVAVFALQAGRIAPLSPVGGVYTD
RSDVLSDSSPAVLTITSSAVDYVQVAVRARGSPSPSITFVDLLDPAQYFKED
EYSPATILQITSGSTRTPAGVYMSHONRVAFEQMGCYFADTDCITPPINSLWMLP
FYHDMGLVIGICAPILGYPALVITSPVSLQRPARMHMLASDFHAEFANAFELA
AARTTDDMAGRDGNILITLISSEBVOAATIKRFADRFARFNLOERYIRSYGLAE
TVVATSKDPGPPEYDFDTELSAGAKPCAGGATSLISYMLRSPYIVISGIDTC
IECPDGYGEIWMGVNDVANGYWMQKPDSESTFGKITYPSPGPEGPNLRTGSGFY
TDCMKFIIRIKDILLIYGRNHSPDIEATTOETTRGCAISVPGDSESTKIVAIT
LKRKGDSDQDAMARLGAIKREVTSALSSHGSLVADLVVAPGSTIPITSGKVRGAC
VEQYRQDQFARLDA"
/gene="WT3012"
/802..11564
/802..11564
/gene="WT3012"
/note="similar to PID:699243 SP:P54881; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, Mmpl family"
/protein_id="AAK4739.1"
/db_xref="GI:13882795"
/translation="MPSPAGRLHRIYIRLKSSPDCRATTSSGADQQRSPRLTNL
LVNAWAAVAIYANILLFTQAEPHDTSPLLQDAAKTAATSRAQAFPTGSNAIA
YLVEGSGTLEPQDQPYDAVAGLRADTRHGSVLDWMSDPVYAPLCITSPDGSATA
MWMREGSGTQOAESLDVARSVLRQLPSEGLRASIYPAITNDMPQIATMOSATI
VTVAIVAVILLRLARLSVRAAYILITADISLAVAMPILAVGVGHMGDSVPSWTL
AAVITGITATATMLAARLGSDAHSNAIPTRDSIPAFALPCVATITGPLLARTP
ALHGVTGRGLGEFVALAASLTIVLPALIALAASQQLPAPTTGAGWTGSLSPVSSASA
LGTAVALAICMLPTIGMMGVANENPTROGGAQVLPGNALPVVVYIKSRDSDRPAALI
AINQVSHRIVEPCKRVESAAMPAGVWPTGASVLSAAGRIADLQOAGSGFVPAVTA
IKSMKSTIEOBSGADVOLDSTVNTLAGRAAOQYLDMPAAARNLKKTTELSYLE
TITHVIIGFTNCDDVILCTAMRKVIEPDIYVYTMNELSTADRSATISOTGNSLSS
APRWAOQRSALAOYRSFVPKLETTIDPAMQIOKASAMKLNLSADPDTGEGHLS
RKDLADPSYRIVRESMFSDDGTATRLFLYSQGLDLAAAAQOLEIAAGAMVYGL
VDSQVTVGGAQOIAAAVSDALIHDAVLAVLILLVVALASMRGSRHAAVAGVYLAS
YLAAGVSIALMOHLIDRELNALVPLVFAVILASGVYIVAGIKAGIADAEATGARS
KGAVSGGAVAPFLAALGSGVAGLVIVSGGSFVLSQIGTVVVLGLVLLIVQARMLP
TTPGR"
/complement(11518..11733)
/gene="WT3013"
/complement(11518..11733)
/gene="WT3013"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47340.1"
/db_xref="GI:13882796"
/translation="MGSYEPRIORPRLPRATARRPLTRPTGPNPYWCLRWYIV
GQPSVANGLEDAVNAALAWSEATLAAP"
/complement(11758..11928)
/gene="WT3014"
/complement(11758..11928)
/gene="WT3014"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47341.1"
/db_xref="GI:13882797"
/translation="MTVAGDGDGLVLGSAMCDVASFNGAVAQRELRARENRRISPSY
HARRFNFTNST"
/gene="WT3015"
/12196..13437
/gene="WT3015"
/12196..13437
/gene="WT3015"
/12196..13437
/gene="WT3015"
```

```
/note="similar to GB:AL123456; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="IS153, OrfA"
/protein_id="AAK47342.1"
/db_xref="GI:13882798"
/translation="MLTVEEDAEIRLHRABGLPIKMIARVILGISKMTVKSALSNQO
PKYERAPGSIYDVAVERIRRLLOAYPTMPATVIAERIGWERSITRYLSARVAELRPY
LPDPASRTTIVAGEIACDDFPPELIPVFGQRTAKQLPLTMCAISRWILLAML
LPSCAEDELFEAGWRLIEALGAVPRVLVWDEGAIGRWGSRSELTQEQKRGTLAA
KVLICRPADPEPAKGLIERAHDLYERSELPGRVFASPADFNQGLAMALVNTRRRL
GCAPTRDIGADRAMLSLPVPAPATGCTSLRLPRDHYRCDSDYSVHPGVIQHRVL
VRADIERVHDCDELVAHDHRIAMAHQTVSDPAHVEAKVLRRRHFSAAAPVVEPOY
OVRSLSDVDAIDGYDIDGVA"
/gene="WT3016"
/13522..14238
/13522..14238
/gene="WT3016"
/note="similar to GB:AL123456; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="IS153, OrfB"
/protein_id="AAK47343.1"
/db_xref="GI:13882799"
/translation="MSQCRGWPPIAPRPTGATKTMWPPACSGKQOPGSPVYVRAASAP
HASRDSRWKSSITLSMLVANSATPSHIAWPISSPATTSCFNAIPGIGKTHLVGLA
IRACQAGRVLFAATAAEVVARLAHNAHGRUYAELTRLCRPLLVVDEVGITPEPEA
ANLFQOLVSSRYERASLIYVSNKAFGRWGEVGGDDVYAAAMIDRLVHBAEVVALKGD
SYRLKRDLDGRVPPAGTTEE"
BASE COUNT 2504 a 4772 c 4554 g 2496 t
ORIGIN
Query Match 50.9%; Score 51.4; DB 1; Length 14326;
Best Local Similarity 69.3%; Pred. No. 0.0093;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 ATCGCCCGCCACTCGCGCCGAGATATTGGACCCGTGTRACCCCGAGAAAGCGGCG 60
11111111 111111111111111111111111111111111111111111111
DB 1716 ATTGCCCGCGCCAGAGGAGCGGATTTTCGCGCGGCAATCCCGAGGAGCGC 1657
61 TCGCTGCGGAGCAGGAGATCGGACGCGATGAGGACTCGCG 101
11111111 111111111111111111111111111111111111111111111
QY 61 TCGCTGCGGAGCAGGAGATCGGACGCGATGAGGACTCGCG 101
11111111 111111111111111111111111111111111111111111111
DB 1656 ATGCTGCGGACATGCGGCTCGAGCATGTCTAGATTGCGC 1616
```

Search completed: November 6, 2002, 13:27:29
Job time : 601.625 secs


```

BASE COUNT      115 a      95 c      153 g      79 t      6 others
ORIGIN

Query Match      33.5%; Score 33.8; DB 14; Length 448;
Best Local Similarity 60.2%; Pred. No. 34;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 9 GGATGATACCGCGGTGGCGGTCAGTCATGAGGACGAGACGACCGCCCGCA 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 149 GAAGGAGACCGCGCGGTGGCGGAGGAGGAGGTGTGGAGGTGGCCCGCGCATGCT 208
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 69 ATTTCAGCTGGCAGCCGCGGTGGCGGCGCACGG 101
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 GGTGCAGAACGCGGAGCGCGAGCTCGAGCCACCG 241
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
CN50209E
LOCUS
DEFINITION Tetradodon nigroviridis genome survey sequence PUC-ori end of clone
157114 of library G from Tetradodon nigroviridis, genomic survey
sequence.
ACCESSION AL209003.1 GI:7867822
VERSION AL209003.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradodon nigroviridis.
ORGANISM Tetradodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetradodon.
1 (bases 1 to 1108)
Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetradodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 1108)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradodon nigroviridis
Unpublished
3 (bases 1 to 1108)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradodon.
Location/Qualifiers
1..1108
/organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone_11b="G"
/clone_11b="G"
/note="Genoscope sequence ID : COAG157BE07SP1-end :
PUC-ori"

BASE COUNT      262 a      255 c      309 g      273 t      9 others
ORIGIN

Query Match      32.1%; Score 32.4; DB 17; Length 1108;
Best Local Similarity 58.7%; Pred. No. 98;
Matches 54; Conservative 1; Mismatches 37; Indels 0; Gaps 0;

OY 10 GATGATACCGCGGTGGCGGTCATGAGGACGAGGAGCGACCGCCCGTCGCA 69
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 731 GATGAGACCMKAGAGACGCGGCGACCGTGKGGCGTGTCTCTCCGGGCGACTCCAC 790
OY 70 TTCACAGTTCGAGCGCGGTGGCGGCGCACGG 101
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 791 CTCACCTGAGAGGACGAGGTATAGGACCCCTG 822
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
BF292924/c
LOCUS
DEFINITION whole plant cDNA library Trilicium turgidum L. var. durum (durum wheat)
WHE2166_E01_J02, mRNA sequence.
WHE2166_E01_J02, mRNA sequence.
ACCESSION BF292924.1 GI:11223988
VERSION BF292924
KEYWORDS EST.
SOURCE Trilicium turgidum.
ORGANISM Trilicium turgidum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Trilicium.
1 (bases 1 to 688)
Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,R., Tong,J.C.
and Zhang,D.
The structure and function of the expressed portion of the wheat
genomes - Whole plant cDNA library from Trilicium turgidum L. var.
durum
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105858773
Fax: 5105858818
Email: oanderson@w.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.
Location/Qualifiers
1..688
/organism="Trilicium turgidum"
/cultivar="Langdon-16"
/db_xref="taxon:4571"
/clone="WHE2166_E01_J02"
/clone_11b="Trilicium turgidum L. var. durum (durum wheat)
whole plant cDNA library"
/tissue_type="All tissues"
/dev_stage="Different growth stages"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: ECORI, Site_2: XhoI; Plants were grown in a growth
chamber at North Dakota State University (Kianian, Otto,
Simons). Tissues collected from seven-day etiolated
seedling leaf, stem, root and seed; leaf from plant at
fourth leaf stage; spike from pre-anthesis through 20
days after anthesis; flag leaf; leaf and stem tissue from
tillers; and root. Total RNA and poly(A) RNA were
prepared from each tissue and then pooled, a cDNA library
was made, and the cDNA clones were in vivo excised to give
pluscript phagemids in the T7 Close lab (Akhunov, Chin,
Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
```

```

BASE COUNT      98 a      301 c      160 g      129 t
ORIGIN

Query Match      31.9%; Score 32.2; DB 12; Length 688;
Best Local Similarity 63.6%; Pred. No. 99;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```


8 CGATGATACGCCGGTGGCGGTGCAGGTCATGGCGACCGAGGAGCGACCGAGGCCGCTGC 67

[illegible]

JOURNAL	COMMENT
Unpublished (2000)	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanderson@w.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: StrataGene SK primer.
FEATURES	Location/Qualifiers
SOURCE	1..489
	/organism="Triticum aestivum"
	/cultivar="Chinese Spring"
	/db_xref="taxon:4565"
	/clone="WHE3455_F02_K03"
	/clone_1lb="Wheat pre-anthesis spike cDNA library"
	/issue_type="spike before anthesis"
	/dev_stage="Adult plant"
	/lab_host="E. coli SOLR"
	/note="Vector: Lambda uni-zap XR, excised phagemid; site_1: EcoRI; site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give plasmid phagemids in the TJ Clone Lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT	68 a 199 c 161 g 61 t
ORIGIN	
Query Match	31.1%; Score 31.4; DB 13; Length 489;
Best Local Similarity	67.7%; Pred. No. 1.5e+02;
Matches 44; Conservative	0; Mismatches 21; Indels 0; Gaps 0;
QY 2	CCCTTCGGCGATGATACGCCGGTGGCGATGCGACGACGAGCGACCGCC 61
DB 174	CTTCGGGCGAGGCGCGCGGTGAGTGCAGCGACGACGAGGCGGACCGT 115
QY 62	GCCG 66
DB 114	GCATG 110
RESULT 8	
LOCUS	B0839382 596 bp mRNA linear EST 08-AUG-2002
DEFINITION	WHE415_E02_J03S Wheat CS whole plant cDNA library Triticum aestivum CDNA clone WHE415_E02_J03, mRNA sequence.
ACCESSION	B0839382
VERSION	B0839382.1 GI:22143704
KEYWORDS	EST.
SOURCE	wheat.
ORGANISM	Triticum aestivum
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
AUTHORS	1 (bases 1 to 596)
TITLE	Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J., Lazo,G.R., Rausch,C.J., Wilson,C. and Woo,J.
JOURNAL	The structure and function of the expressed portion of the wheat genomes - Chinese Spring whole plant cDNA library
COMMENT	Unpublished (2002)
	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: oanderson@w.usda.gov

DEFINITION 1700068749053 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 ACCESSION 19600449683838 5', mRNA sequence.
 VERSION BM597575
 KEYWORDS BM597575.1 GI:18895678
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 655)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004825 row: G column: 24
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..655
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449683838"
 /clone_1id="A.Gam.ad.cDNA.blood1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
 hours after human blood feeding. cDNA inserts >500 bp
 cloned directionally into pSport 1. Not 1 site is 3'.
 Clones available through the Malaria Research and
 Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 119 a 187 c 233 g 116 t
 ORIGIN
 Query Match 30.7%; Score 31; DB 13; Length 655;
 Best Local Similarity 64.8%; Pred. No. 2e+02;
 Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 9 GGATGATACGCCGGTGGCGGTGATGTCAGACGACGAGGAGGACCGCCGCTGCA 68
 || ||| |||| |||| ||| || || || |||| |||| ||| ||
 Db 350 GGGGATCAGCGCGCTGGCGCTTCAACCGCTGCACGCGGAGCTTGTCCGACACCC 291
 || ||| |||| |||| ||| || || || |||| |||| ||| ||
 QY 69 ATTCCACGTTG 79
 || ||| ||||
 Db 290 ACTCCTCGTTG 280
 || ||| ||||
 RESULT 14
 BM589723/c 681 bp mRNA linear EST 25-FEB-2002
 LOCUS 1700068749053 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 DEFINITION 19600448421008 5', mRNA sequence.
 ACCESSION BM589723
 VERSION BM589723
 KEYWORDS BM589723.1 GI:18885584
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 681)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004825 row: E column: 16
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..687
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449702982"
 /clone_1id="A.Gam.ad.cDNA.blood1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
 hours after human blood feeding. cDNA inserts >500 bp
 cloned directionally into pSport 1. Not 1 site is 3'.
 Clones available through the Malaria Research and
 Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 142 a 198 c 221 g 120 t
 ORIGIN
 Query Match 30.7%; Score 31; DB 13; Length 681;
 Best Local Similarity 64.8%; Pred. No. 2e+02;
 Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 9 GGATGATACGCCGGTGGCGGTGATGTCAGACGACGAGGAGGACCGCCGCTGCA 68
 || ||| |||| |||| ||| || || || |||| |||| ||| ||
 Db 116 GGGGATCAGCGCGCTGGCGCTTCAACCGCTGCACGCGGAGCTTGTCCGACACCC 57
 || ||| |||| |||| ||| || || || |||| |||| ||| ||
 QY 69 ATTCCACGTTG 79
 || ||| ||||
 Db 56 ACTCCTCGTTG 46
 || ||| ||||
 RESULT 15
 BM613694/c 687 bp mRNA linear EST 25-FEB-2002
 LOCUS 17000687139208 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 DEFINITION 19600449702982 5', mRNA sequence.
 ACCESSION BM613694
 VERSION BM613694
 KEYWORDS BM613694.1 GI:18911800
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 687)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004825 row: E column: 16
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..687
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449702982"

COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004911 row: F column: 07
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..681
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600448421008"
 /clone_1id="A.Gam.ad.cDNA.blood1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
 hours after human blood feeding. cDNA inserts >500 bp
 cloned directionally into pSport 1. Not 1 site is 3'.
 Clones available through the Malaria Research and
 Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 142 a 198 c 221 g 120 t
 ORIGIN
 Query Match 30.7%; Score 31; DB 13; Length 681;
 Best Local Similarity 64.8%; Pred. No. 2e+02;
 Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 9 GGATGATACGCCGGTGGCGGTGATGTCAGACGACGAGGAGGACCGCCGCTGCA 68
 || ||| |||| |||| ||| || || || |||| |||| ||| ||
 Db 116 GGGGATCAGCGCGCTGGCGCTTCAACCGCTGCACGCGGAGCTTGTCCGACACCC 57
 || ||| |||| |||| ||| || || || |||| |||| ||| ||
 QY 69 ATTCCACGTTG 79
 || ||| ||||
 Db 56 ACTCCTCGTTG 46
 || ||| ||||
 RESULT 15
 BM613694/c 687 bp mRNA linear EST 25-FEB-2002
 LOCUS 17000687139208 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
 DEFINITION 19600449702982 5', mRNA sequence.
 ACCESSION BM613694
 VERSION BM613694
 KEYWORDS BM613694.1 GI:18911800
 SOURCE EST.
 ORGANISM African malaria mosquito.
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 REFERENCE 1 (bases 1 to 687)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 'R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU01004825 row: E column: 16
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..687
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449702982"

```
/clone_11b="A.Gam.ad.cDNA.blood1"  
/dev_stage="Adult"  
/lab_host="DH10b"  
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
hours after human blood feeding. cDNA inserts >500 bp  
cloned directionally into pSport 1. Not I site is 3'.  
Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)"
```

```
BASE COUNT      145 a      195 c      224 g      123 t  
ORIGIN
```

```
Query Match      30.7%; Score 31; DB 13; Length 687;  
Best Local Similarity 64.8%; Pred. No. 2e+02;  
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
```

```
QY      9  GGATGATACGCGCGGTGGCGGTCAGTCATGCGGACCGAGAGGACGACCGCGCTGCA 68  
        || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db      164 GGCGGATCAGCGCGCTGCGCTCAACCCGTCGACGCGGAAGCGGACTTGTCCGACAGCC 105  
QY      69  ATTCAACGTTG 79  
        || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db     104 ACTCTCTGTTG 94
```

```
Search completed: November 6, 2002, 15:52:40  
Job time : 1201.62 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 ; Search time 27.125 Seconds
(without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_34350_34450

Perfect score: 101

Sequence: 1 gcttcgcgcatgatacgc.....agccgggtgcggcgccacgg 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 16577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCr_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	29.7	167343	10	US-09-962-436-281
2	30	29.7	167343	10	US-09-964-824A-273
3	28.6	28.3	3039	10	US-09-792-127-3
4	28.4	28.1	4127	10	US-09-854-883-10
5	28.4	27.1	324	10	US-09-917-800A-1555
6	27.4	27.1	324	10	US-09-920-300A-576
7	27.4	27.1	324	10	US-10-033-528-576
8	27.4	27.1	2162	10	US-09-799-875-10
9	27.2	26.9	939	10	US-09-021-660A-33
10	27.2	26.9	1622	8	US-08-900-220C-7
11	27.2	26.9	1622	10	US-09-151-999-7
12	27.2	26.9	1985	10	US-09-925-302-332
13	27.2	26.9	2108	10	US-09-954-456-686
14	27.2	26.9	2108	10	US-09-954-456-1117
15	27.2	26.9	2108	10	US-09-880-107-1673
16	27.2	26.9	11220	10	US-09-861-289-32
17	27.2	26.9	36778	10	US-09-861-289-5
18	27	26.7	1455	10	US-09-815-242-7924
19	26.8	26.5	2100	10	US-09-948-777-1

c	20	26.6	26.3	298	10	US-09-294-093B-4898	Sequence 4898, Ap
	21	26.6	26.3	603	10	US-09-893-737-267	Sequence 267, App
	22	26.6	26.3	1101	10	US-09-438-623A-3	Sequence 3, App11
c	23	26.2	25.9	315	10	US-09-867-701-4939	Sequence 4939, Ap
	24	26.2	25.9	475	10	US-09-864-761-6200	Sequence 6200, Ap
c	25	26.2	25.9	510	10	US-09-867-701-3223	Sequence 3223, Ap
	26	26.2	25.9	111282	12	US-10-094-989-3	Sequence 3, App11
c	27	26.2	25.7	425	10	US-09-864-761-690	Sequence 690, App
	28	26	25.7	2027	10	US-09-867-550-887	Sequence 887, App
c	29	26	25.7	3236	10	US-09-954-456-43	Sequence 43, App1
	30	26	25.7	3936	10	US-09-919-172-49	Sequence 49, App1
	31	25.8	25.5	453	10	US-09-887-576-797	Sequence 797, App
	32	25.8	25.5	847	10	US-09-833-381-44	Sequence 44, App1
c	33	25.8	25.5	2223	10	US-09-897-114-13	Sequence 13, App1
	34	25.6	25.3	240	10	US-09-960-352-1820	Sequence 1820, Ap
	35	25.6	25.3	411	10	US-09-960-352-5646	Sequence 5646, Ap
	36	25.6	25.3	419	10	US-09-960-352-8465	Sequence 8465, Ap
	37	25.6	25.3	427	10	US-09-960-352-8465	Sequence 146, App
	38	25.6	25.3	1891	10	US-09-969-708-146	Sequence 6637, Ap
	39	25.4	25.1	413	10	US-09-960-352-6637	Sequence 7705, Ap
	40	25.4	25.1	918	10	US-09-815-242-7705	Sequence 1, App11
c	41	25.4	25.1	1077	9	US-10-104-340-1	Sequence 363, App
	42	25.4	25.1	1324	10	US-09-925-301-363	Sequence 363, App
c	43	25.4	25.1	1324	10	US-09-925-301-363	Sequence 56, App1
	44	25.4	25.1	1521	12	US-10-044-090-56	Sequence 8009, Ap
	45	25.4	25.1	1788	10	US-09-815-242-8009	

ALIGNMENTS

RESULT 1

US-09-962-436-281

Sequence 281, Application US/09962436

Patent No. US20020081301A1

GENERAL INFORMATION:

APPLICANT: Soppet, Daniel

TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign

FILE REFERENCE: 689290-75

CURRENT APPLICATION NUMBER: US/09/962,436

PRIOR FILING DATE: 2001-09-25

PRIOR APPLICATION NUMBER: US/60/235,082

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/234,924

PRIOR FILING DATE: 2000-09-25

NUMBER OF SEQ ID NOS: 568

SOFTWARE: PatentIn version 3.0

SEQ ID NO 281

LENGTH: 167343

TYPE: DNA

ORGANISM: Homo sapiens

US-09-962-436-281

Query Match 29.7%; Score 30; DB 10; Length 167343;

Best Local Similarity 67.7%; Pred. No. 1.9;

Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 40 GCGACGAGAGCGACGACGCGCTCCATTCACGTTGCGAGCGGTCGCGGCGAC 99

Db 117079 GCGTCGAGAACGCGCGCGCTCCGACGCTCCGCGGTCGACACCGGTCAGGGGCGCG 117138

QY 100 GG 101

Db 117139 GG 117140

RESULT 2

US-09-964-824A-273

Sequence 273, Application US/09964824A

Patent No. US20020102531A1

GENERAL INFORMATION:

APPLICANT: Horrigan, Stephen

```
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 273
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-273
```

```
Query Match 29.7%; Score 30; DB 10; Length 167343;
Best Local Similarity 67.7%; Pred. No. 1.9;
Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
```

```
OY 40 GCGACCGAGGAGCGACCGCCCTGCAATTCGAGTTGCGAGCGGGTCCGGCCAC 99
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117079 GCGTCGGAAGACCGCGAGCGCTCCGCACTCCGCGGTGACACCGGGGTGAGGGCGC 117138
OY 100 GG 101
Db 117139 GG 117140
```

```
RESULT 3
US-09-792-127-3/C
; Sequence 3, Application US/09792127
; Patent No. US20020002713A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Beckles, Diane M.
; APPLICANT: Butler, Karla
; APPLICANT: Pearlstein, Rich
; TITLE OF INVENTION: Starch Branching Enzyme IIb
; FILE REFERENCE: BBI439 US NA
; CURRENT APPLICATION NUMBER: US/09/792,127
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/186098
; PRIOR FILING DATE: 2000-03-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 3
; LENGTH: 3039
; TYPE: DNA
; ORGANISM: Trillium aestivum
US-09-792-127-3
```

```
Query Match 28.3%; Score 28.6; DB 10; Length 3039;
Best Local Similarity 55.6%; Pred. No. 3.7;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
```

```
OY 2 CCTTCGGGATGATAGCCGGTGGCGGTGAGTTCATGGCGAGGAGGAGGACGAGGCC 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 CCGACGCCGAGCGGCGGCGGAGCGGGTGAACGGGTGCCCTTGTGGCGGCAAGACGATGGC 164
OY 62 GCCTGCAATTCAGGTTGGAGCGCGGGTGCAGCGGGCCACG 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 GACTGCACTCCACCCCGGCGCCCTCCGCTGCGCCCG 125
```

```
RESULT 4
US-09-854-883-10
; Sequence 10, Application US/09854883
; Patent No. US20020055479A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lex M. Cowser
; APPLICANT: Jacqueline Wyalt
; APPLICANT: Susan M. Freier
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Robert McKay
; TITLE OF INVENTION: ANTISENSE MODULATION OF PTP1B EXPRESSION
; FILE REFERENCE: ISPH-0576
; CURRENT APPLICATION NUMBER: US/09/854,883
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/629,644
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/487,368
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 389
; SEQ ID NO: 10
; LENGTH: 4127
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(1418)
US-09-854-883-10
```

```
Query Match 28.1%; Score 28.4; DB 10; Length 4127;
Best Local Similarity 62.9%; Pred. No. 4.4;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
```

```
OY 18 GCGGTGGCGGCTGATGCGAGCGAGCGAGGAGCGAGCGCGGCTCATTCACCT 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 GCAGTTGAGTTGAGACACTGGCTACCCAGAGGCTCGAGAGATCTGCAATTCACCTA 646
OY 78 TCGAGCCCG 87
Db 647 CACCACCTGG 656
```

```
RESULT 5
US-09-917-800A-1555
; Sequence 1555, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 1555
```



```
; LENGTH: 4127
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_012637
US-09-917-800A-1555

Query Match
Best Local Similarity 28.1%; Score 28.4; DB 10; Length 4127;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 18 GCCGGTGGCGGTGAGTGCATGCGACCGAGAGCGACCGCCCTGCAATTCACGG 77
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 587 GCAGTTGAGTTGGAGAACCTGCTACCCAGAGAGCGCTGAGAGATCCGATTCCTCACTA 646
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 78 TCGAGCCGG 87
    | | | | |
Db 647 CACCACCTGG 656

RESULT 6
US-09-920-300A-576
; Sequence 576, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 576
; LENGTH: 324
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-576

Query Match
Best Local Similarity 27.1%; Score 27.4; DB 10; Length 324;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 9 GGATGATACGCCGGTGGCGGTGACAGTCAATGCGACCGAGAGCGACCGCCGCTGCA 68
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 GGATGATGTTGGCGGTGACTCTCCAGCGGAGCGGCCCGACCTTGGCGGCAATCCTGT 269
    | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 ATTCCAGCTTCCGAGCCGGGTGCGGCGCCACGG 101
    | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270 AGGCGAGGTGAAGCCCGCTGACGATCTCAG 302

RESULT 7
US-10-033-528-576
; Sequence 576, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1886
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 576
; LENGTH: 324
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-576

Query Match
Best Local Similarity 27.1%; Score 27.4; DB 12; Length 324;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 9 GGATGATACGCCGGTGGCGGTGACAGTCAATGCGACCGAGAGCGACCGCCGCTGCA 68
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 GGATGATGTTGGCGGTGACTCTCCAGCGGAGCGGCCCGACCTTGGCGGCAATCCTGT 269
    | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 ATTCCAGCTTCCGAGCCGGGTGCGGCGCCACGG 101
    | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270 AGGCGAGGTGAAGCCCGCTGACGATCTCAG 302

RESULT 8
US-09-799-875-10
; Sequence 10, Application US/09799875
; Patent No. US20020034780A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: No. US20020034780A1 Human Protein Kinases and Uses
; FILE REFERENCE: 35800/209996
; CURRENT APPLICATION NUMBER: US/09/799,875
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/182,059
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/659,287
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 2162
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (76)...(1818)
US-09-799-875-10

Query Match
Best Local Similarity 27.1%; Score 27.4; DB 10; Length 2162;
Matches 55; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCCCGGTGGCGGTGACAGTCAATGCGGACCGAGAGCGACCG 60
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 GCGTCCGCCACGCGCTCCGCGCGGTGGTGGCGGAGCGGCGGTGGCGGACCGGCG 63
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY 61 CGCTGCAATTCACAGTTCGAGCGCGGCGGTGCGGCGCCACGG 101
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 64 CGGCGCGCCACCATGCGCGGTGCGAGACGCGCTGGGCGCGG 104

RESULT 9
US-09-021-660A-33
; Sequence 33, Application US/09021660A
; Patent No. US20010041668A1
; GENERAL INFORMATION:
; APPLICANT: Baron, M.
; APPLICANT: Farrington, S.
; APPLICANT: Belaussoff, M.
; TITLE OF INVENTION: METHODS FOR MODULATING HEMATOPOIESIS AND VASCULAR
; FILE REFERENCE: H01P-P01-060
; CURRENT APPLICATION NUMBER: US/09/021,660A
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 60/037,513
; PRIOR FILING DATE: 1997-02-10
```

PRIOR APPLICATION NUMBER: 60/049,763
PRIOR FILING DATE: 1997-06-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 33
LENGTH: 939
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(939)
US-09-021-660A-33

Query Match
Best Local Similarity 26.9%; Score 27.2; DB 10; Length 939;
61.1%; Pred. No. 8.8;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCAGTATGCGACCGAGAGCGACCGC 60
DB 445 GACCGGAGCGCCCGACAGGCTGAGAGCCTTCCAGTATGAGACTGAGACCGCCCGACGC 504
QY 61 CGCCTGCAATTC 72
DB 505 CGCCTGGCACTC 516

RESULT 10
US-08-900-220C-7
Sequence 7, Application US/08900220C
Patent No. US20020045206A1
GENERAL INFORMATION:
APPLICANT: Miao, Ningning
Mang, Monica
Mahanchappa, Nagesh K.
Jin, Ping
Jin, Ping
TITLE OF INVENTION: Method of Treating Dopaminergic and
GABA-nergic Disorders
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: ONE POST OFFICE SQUARE
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,220C
FILING DATE: 24-Jul-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: ONV-044.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1622 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 51..1283
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-08-900-220C-7
Query Match
Best Local Similarity 26.9%; Score 27.2; DB 8; Length 1622;
61.1%; Pred. No. 9.2;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCAGTATGCGACCGAGAGCGACCGC 60
DB 792 GACCGGAGCGCCCGACAGGCTGAGAGCCTTCCAGTATGAGACTGAGACCGCCCGACGC 851
QY 61 CGCCTGCAATTC 72
DB 852 CGCCTGGCACTC 863

RESULT 11
US-09-151-999-7
Sequence 7, Application US/09151999
Patent No. US20020151460A1
GENERAL INFORMATION:
APPLICANT: Wang, Elizabeth
TITLE OF INVENTION: REGULATION OF EPITHELIAL TISSUE BY HEDGEHOG-LIKE
POLYPEPTIDES, AND FORMULATIONS AND USES RELATED THERETO
FILE REFERENCE: ONV-031.02
CURRENT FILING DATE: 1998-08-11
EARLIER APPLICATION NUMBER: 08/955,552
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 7
LENGTH: 1622
TYPE: DNA
ORGANISM: Homo sapien Ihh
FEATURE:
NAME/KEY: CDS
LOCATION: (51)..(1283)
US-09-151-999-7

Query Match
Best Local Similarity 26.9%; Score 27.2; DB 10; Length 1622;
61.1%; Pred. No. 9.2;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCAGTATGCGACCGAGAGCGACCGC 60
DB 792 GACCGGAGCGCCCGACAGGCTGAGAGCCTTCCAGTATGAGACTGAGACCGCCCGACGC 851
QY 61 CGCCTGCAATTC 72
DB 852 CGCCTGGCACTC 863

RESULT 12
US-09-925-302-332/C
Sequence 332, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antbodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 332
LENGTH: 1985
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

```

: NAME/KEY: misc feature
: LOCATION: (360)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (1985)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-332
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 1985;
Best Local Similarity 56.8%; Pred. No. 9.3;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 12 TGATACGGCGGTGGCGGTGATGTCATGCGACCGAGGAGCGCGCCCTGCATTT 71
    || || || || || || || || || || || || || || || || || || ||
Db 1482 TGTGAGCTCTCTGGCGGACGAGGTGTAGAGCGGACGCTGCGTGGAGCCACCA 1423
```

```
QY 72 CCAGTTCGAGCGCGGTGGCGGCGCCAC 99
    | | | | | | | | | | | | | | | |
Db 1422 GGAAGGCGCGGCGAGGTGTGAAGGCCAC 1395
```

RESULT 13

```
US-09-954-456-686/c
: Sequence 686, Application US/09954456
: Patent No. US20020115057A1
: GENERAL INFORMATION:
```

```
: APPLICANT: Young, Paul
```

```
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
```

```
: FILE REFERENCE: 689290-76
```

```
: CURRENT APPLICATION NUMBER: US/09/954,456
```

```
: PRIOR FILING DATE: 2001-09-18
```

```
: PRIOR APPLICATION NUMBER: US/60/233,617
```

```
: PRIOR FILING DATE: 2000-09-18
```

```
: PRIOR APPLICATION NUMBER: US/60/234,052
```

```
: PRIOR FILING DATE: 2000-09-20
```

```
: PRIOR APPLICATION NUMBER: US/60/234,923
```

```
: PRIOR FILING DATE: 2000-09-25
```

```
: PRIOR APPLICATION NUMBER: US/60/235,134
```

```
: PRIOR FILING DATE: 2000-09-25
```

```
: PRIOR APPLICATION NUMBER: US/60/235,637
```

```
: PRIOR FILING DATE: 2000-09-26
```

```
: PRIOR APPLICATION NUMBER: US/60/235,638
```

```
: PRIOR FILING DATE: 2000-09-26
```

```
: PRIOR APPLICATION NUMBER: US/60/235,711
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: PRIOR APPLICATION NUMBER: US/60/235,720
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: PRIOR APPLICATION NUMBER: US/60/235,840
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: PRIOR APPLICATION NUMBER: US/60/235,863
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: NUMBER OF SEQ ID NOS: 2276
```

```
: SOFTWARE: PatentIn version 3.0
```

```
: SEQ ID NO 686
```

```
: LENGTH: 2108
```

```
: TYPE: DNA
```

```
: ORGANISM: Homo sapiens
```

```
US-09-954-456-686
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 2108;
```

```
Best Local Similarity 56.8%; Pred. No. 9.3;
```

```
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 12 TGATACGGCGGTGGCGGTGATGTCATGCGACCGAGGAGCGCGCCCTGCATTT 71
```

```
    || || || || || || || || || || || || || || || || || || ||
```

```
Db 1648 TGTGAGCTCTCTGGCGGACGAGGTGTAGAGCGGACGCTGCGTGGAGCCACCA 1589
```

```
QY 72 CCAGTTCGAGCGCGGTGGCGGCGCCAC 99
```

```
    | | | | | | | | | | | | | | | |
```

```
Db 1588 GGAAGGCGCGGCGAGGTGTGAAGGCCAC 1561
```

RESULT 14

```
US-09-954-456-1117/c
: Sequence 1117, Application US/09954456
: Patent No. US20020115057A1
: GENERAL INFORMATION:
```

```
: APPLICANT: Young, Paul
```

```
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
```

```
: FILE REFERENCE: 689290-76
```

```
: CURRENT APPLICATION NUMBER: US/09/954,456
```

```
: PRIOR FILING DATE: 2001-09-18
```

```
: PRIOR APPLICATION NUMBER: US/60/233,617
```

```
: PRIOR FILING DATE: 2000-09-18
```

```
: PRIOR APPLICATION NUMBER: US/60/234,052
```

```
: PRIOR FILING DATE: 2000-09-20
```

```
: PRIOR APPLICATION NUMBER: US/60/234,923
```

```
: PRIOR FILING DATE: 2000-09-25
```

```
: PRIOR APPLICATION NUMBER: US/60/235,134
```

```
: PRIOR FILING DATE: 2000-09-25
```

```
: PRIOR APPLICATION NUMBER: US/60/235,637
```

```
: PRIOR FILING DATE: 2000-09-26
```

```
: PRIOR APPLICATION NUMBER: US/60/235,638
```

```
: PRIOR FILING DATE: 2000-09-26
```

```
: PRIOR APPLICATION NUMBER: US/60/235,711
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: PRIOR APPLICATION NUMBER: US/60/235,720
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: PRIOR APPLICATION NUMBER: US/60/235,840
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: PRIOR APPLICATION NUMBER: US/60/235,863
```

```
: PRIOR FILING DATE: 2000-09-27
```

```
: NUMBER OF SEQ ID NOS: 2276
```

```
: SOFTWARE: PatentIn version 3.0
```

```
: SEQ ID NO 1117
```

```
: LENGTH: 2108
```

```
: TYPE: DNA
```

```
: ORGANISM: Homo sapiens
```

```
US-09-954-456-1117
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 2108;
```

```
Best Local Similarity 56.8%; Pred. No. 9.3;
```

```
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 12 TGATACGGCGGTGGCGGTGATGTCATGCGACCGAGGAGCGCGCCCTGCATTT 71
```

```
    || || || || || || || || || || || || || || || || || || ||
```

```
Db 1648 TGTGAGCTCTCTGGCGGACGAGGTGTAGAGCGGACGCTGCGTGGAGCCACCA 1589
```

```
QY 72 CCAGTTCGAGCGCGGTGGCGGCGCCAC 99
```

```
    | | | | | | | | | | | | | | | |
```

```
Db 1588 GGAAGGCGCGGCGAGGTGTGAAGGCCAC 1561
```

RESULT 15

```
US-09-880-107-1673/c
: Sequence 1673, Application US/09880107
: Patent No. US20020142981A1
: GENERAL INFORMATION:
```

```
: APPLICANT: Horne, Darci T.
```

```
: APPLICANT: Vockley, Joseph G.
```

```
: APPLICANT: Scherf, Uwe
```

```
: TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
```

```
: FILE REFERENCE: 44921-5028-WO
```

```
: CURRENT APPLICATION NUMBER: US/09/880,107
```

```
: PRIOR FILING DATE: 2001-06-14
```

```
: PRIOR APPLICATION NUMBER: US 60/211,379
```

```
: PRIOR FILING DATE: 2000-06-14
```

```
: PRIOR APPLICATION NUMBER: US 60/237,054
```

```
: PRIOR FILING DATE: 2000-10-02
```

```
: NUMBER OF SEQ ID NOS: 3950
```

```
: SOFTWARE: PatentIn Ver. 2.1
```

```
: SEQ ID NO 1673
```

```
: LENGTH: 2108
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D50914
US-09-880-107-1673

```

```

Query Match      26.9%; Score 27.2; DB 10; Length 2108;
Best Local Similarity 56.8%; Pred. No. 9.3;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```

```

QY 12 TGATAGCCGGTGGCGGTGTCAGTTCATGGCGACGAGAGCGACGAGCCCGCTGCATT 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1648 TGGTGAAGCTCTCGCGCAGCAGGTGTAAGCGGAGCGAGCGTGCCTGGGAGCCACCACACA 1589

```

```

QY 72 CCACGTTGCGAGCCGGGTGCGGCCAC 99
    ||||| ||||| ||||| ||||| |||||

```

```

Db 1588 GGAAGGGCGCGCAGGTGGAAGCCAC 1561

```

```

Search completed: November 6, 2002, 20:32:45
Job time : 103.125 secs

```



```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 60
|||||
DB 39962 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 40021
```

```
OY 61 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 101
|||||
```

```
DB 40022 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 40062
```

RESULT 3

```
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 60
|||||
DB 39962 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 40021
```

```
OY 61 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 101
|||||
```

```
DB 40022 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 40062
```

RESULT 4

```
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 60
|||||
DB 39962 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 40021
```

```
OY 61 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 101
|||||
```

```
DB 40022 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 40062
```

RESULT 5

```
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIORITY FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 60
|||||
DB 39962 GCCTGCGGATGATACGCCGGTGGCGGTGCAGTGCATGGCGACCGAGGACGACGAGC 40021
```

```
OY 61 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 101
|||||
```

```
DB 40022 CGCTGCAATTCCACGTTGGAGCCGGGTGCCGGGCCACGG 40062
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 635459
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 60
Db 39962 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 40021

QY 61 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 101
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 40062

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 60
Db 39962 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 40021

QY 61 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 101
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 40062
```

```
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 40062

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.8e-19;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 60
Db 39962 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 40021

QY 61 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 101
Db 40022 GCCTGCAATTCCACGTTGCGAGCCGGGTGCCGGCCACG 40062

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: 24366-20007.00
; CURRENT FILING DATE: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          31.3%; Score 31.6; DB 4; Length 4403765;
Best Local Similarity 56.3%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACCGCGGTGGCGTGCAGTCAATGCGAGGAGGACGACGAGC 60
```

DB 4263132 GCCTCCATTCGATATTCGACGGTGCATCCAGAACAGCCATGCACCTTCTCGCC 4263133
OY 61 CGCCTGCAATTCACGCTGGACGCGGTCCGG 94
DB 4263132 GGCCTCCAGACACATCTTTGATCTCGAATCGG 4263099

RESULT 10

US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 441529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 31.3%; Score 31.6; DB 4; Length 441529;
Best Local Similarity 58.5%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCATGCGCAGCGACGACGACGCG 60
DB 4270935 GCCTCCATTCGATATTCGACGGTGCATCCAGAACAGCCATGCACCTTCTCGCC 4270876
OY 61 CGCCTGCAATTCACGCTGGACGCGGTCCGG 94
DB 4270875 GGCCTCCAGACACATCTTTGATCTCGAATCGG 4270842

RESULT 11

US-08-998-416-622/C
; Sequence 622, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwalls Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416

; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 622:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1152 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1421UP
US-08-998-416-622

Query Match 29.3%; Score 29.6; DB 4; Length 1152;
Best Local Similarity 61.8%; Pred. No. 4.3;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 1 GCCTTGGCGATGATACGCCGGTGGCGGTGCATGCGCAGCGACGACGACGCG 60
DB 704 GGCCTCCAGACACGCGCGGTGGCGGTCTTGCGACGACGACGACGACGCGCG 645
OY 61 CGCCTGCAATTCACGCG 76
DB 644 CTGCTGAAGTCGTCG 629

RESULT 12

US-08-822-586-45
; Sequence 45, Application US/08822586
; Patent No. 6015890
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
; APPLICANT: AMALIO TELENTE
; TITLE OF INVENTION: AN EMBCAB OPERON OF MYCOBACTERIA AND
; TITLE OF INVENTION: MUTANTS THEREOF
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,586
; FILING DATE: MARCH 20, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/437
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10095

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
US-08-822-586-45

Query Match 28.7%; Score 29; DB 3; Length 10095;
Best Local Similarity 55.4%; Pred. No. 7;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 GCCTTCGGATGATACCGCGGTGCGAGTATGCGACCGAGACGACGAGC 60
DB 865 GCGCTCACCGCGCGCGCGCTGTGCTGCACATCTGACACCGCGGCGATGCGG 924
QY 61 GCGCTGCATTCACAGTTCGACCGCGGTGCGCGGCGCGC 101
DB 925 CACCGCGGCTTCCTGCGCGCGCTGTGCTGACCGCGCG 965

RESULT 13
US-09-036-987A-1

Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

QY 6 CGCGATGATACGCGGTGCGGTGCGAGTATGCGACCGAGGAGCGACCGCGCGCT 65
Query Match 28.5%; Score 28.8; DB 3; Length 80161;
Best Local Similarity 58.0%; Pred. No. 8.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

DB 49072 CGTGACGCGACACCTGATGTCGTGCGAGCGCTTCCGCGCGCGTGCAGGAGAAAC 49131
QY 66 GCAATTCACGCTGGGACCGCGGTGCGCG 93
DB 49132 GCAGCTGGCACTTCGCGCGCGGTGCGCG 49159

RESULT 14
US-09-370-700-1

Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT FILING DATE: 1999-08-09
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 28.5%; Score 28.8; DB 4; Length 80161;
Best Local Similarity 58.0%; Pred. No. 8.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 6 CGCGATGATACGCGGTGCGGTGCGAGTATGCGACCGAGGAGCGACCGCGCT 65
DB 49072 CGTGACGCGACACCTGATGTCGTGCGAGCGCTTCCGCGCGCGTGCAGGAGAAAC 49131
QY 66 GCAATTCACGCTGGGACCGCGGTGCGCG 93
DB 49132 GCAGCTGGCACTTCGCGCGCGGTGCGCG 49159

RESULT 15

US-09-487-368A-10
Sequence 10, Application US/09487368A
Patent No. 6261840
GENERAL INFORMATION:
APPLICANT: Lex M. Cowsett
APPLICANT: Jacqueline Wyatt
TITLE OF INVENTION: ANTISENSE MODULATION OF PPI1B EXPRESSION
FILE REFERENCE: RTS-0093
CURRENT APPLICATION NUMBER: US/09/487,368A
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 240
SEQ ID NO 10
LENGTH: 4127
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
NAME/KEY: CDS
LOCATION: (120)...(1418)
US-09-487-368A-10

QY 18 GCGGTGCGGTGCGGTGCGGTGCGAGTATGCGACCGAGGAGCGACCGCGCGCTGCAATTCACGCT 77
Query Match 28.1%; Score 28.4; DB 4; Length 4127;
Best Local Similarity 62.9%; Pred. No. 9.8;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db	587	GCAGTTGGAGTTGGAGAACCTGGCTACCCAGAGGCTCGAGAGATCCTGCATTTCACACTA	646
QY	78	TGGAGCCGG	87
Db	647	CACCACCTGG	656

Search completed: November 6, 2002, 16:56:55
Job time : 1888.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 157 Seconds
(without alignments)
1448.738 Million cell updates/sec

Title: US-09-724-876-2_COPY_34350_34450

Perfect score: 101

Sequence: 1 gcctcgcgatgatacgcc.....agccgggtgcccggccacag 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	71989	21	AAA29349	Sorangium cellulos
2	99.4	98.4	68750	21	AAZ55887	Sorangium cellulos
3	31.6	31.3	4403765	22	AA199683	Mycobacterium tube
4	31.6	31.3	4411529	22	AA199682	Mycobacterium tube
5	30.4	30.1	947	23	AB106373	Drosophila melanog
6	30.4	30.1	2890	23	AB106348	Drosophila melanog
7	30.4	30.1	2947	23	AB106372	Drosophila melanog
8	30.4	30.1	4282	23	AB108162	Drosophila melanog
9	30	29.7	167343	24	AB164403	Stomach cancer rel

	10	30	29.7	167343	24	ABL67239	Thyroid cancer rel
	C	11	29.4	1620	21	AAZ53393	Neisseria meningit
	12	29	28.7	10095	19	AAV58938	Mycobacterium tube
	13	28.8	28.5	14775	22	AAE88338	S. spirosa DNA fra
	14	28.8	28.5	50000	22	AAE88313	S. spirosa DNA fra
	15	28.8	28.5	50000	22	AAE88316	S. spirosa DNA fra
	16	28.8	28.5	80161	20	AAZ21501	DNA fragment of Sa
C	17	28.6	28.3	2968	22	AAH78342	Nucleotide sequenc
C	18	28.6	28.3	3039	24	ABK15494	Wheat starch branc
	19	28.4	28.1	1634	24	AB199685	Mouse ischaemic co
	20	28.4	28.1	4127	22	AAD11953	Rat protein phosph
	21	28.4	28.1	4127	24	ABK85028	Rat CDNA for prote
	22	28.4	28.1	4127	24	ABK63648	Rat sequence diffe
	23	28.4	28.1	4127	24	ABK37197	Rat protein tyrosi
C	24	28.2	27.9	2721	12	AAQ10212	BamH1 J-I fragment
C	25	28.2	27.9	2721	12	AAQ10543	BamH1 J-I fragment
C	26	28.2	27.9	3446	23	AB150557	Micromonospora car
	27	28.2	27.9	4583	21	AAA81541	N. meningitidis pa
C	28	28.2	27.9	8438	15	AAO73500	DNA encoding pseud
C	29	28.2	27.9	27541	22	AAD17185	Streptomyces nous
C	30	28.2	27.9	65140	22	AAD17184	Streptomyces nous
C	31	28.2	27.9	125401	22	AAD17186	Streptomyces nous
	32	28.2	27.9	349980	21	AAE21608	Neisseria meningit
	33	28	27.7	990	21	AAA47168	DNA encoding a Cal
	34	28	27.7	28804	17	AAT37329	Sphingon biosynthe
C	35	28	27.7	28804	18	AAE92474	Sphingomonas genu
	36	28	27.7	28804	20	AAV99812	Chromosomal fragme
	37	28	27.7	28804	20	AAV81474	Chromosomal fragme
	38	27.8	27.5	9431	13	AAO20907	DNA encoding the c
C	39	27.6	27.3	2519	23	AB104383	Drosophila melanog
40	27.6	27.3	4985	23	AB104382	Drosophila melanog	
	41	27.6	27.3	35026	21	AAE44890	Bordetella pertuss
42	27.4	27.1	324	24	ABK45025	CDNA encoding colo	
43	27.4	27.1	523	21	AAE35675	Arabidopsis thalids	
C	44	27.4	681	23	AAE67650	DNA encoding novel	
45	27.4	27.1	700	23	AAE6088	DNA encoding novel	

ALIGNMENTS

RESULT 1	AAA29349	standard; DNA; 71989 BP.
ID	AAA29349	
AC	AAA29349;	
XX	12-SEP-2000	(first entry)
DE	Sorangium cellulosum	epothilone polyketide synthase operon genomic DNA.
XX		
KW	Epithilone; polyketide synthase; epocA; epob; epoc; epod; epoe; epof;	
KW	epoL; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;	
KW	tubulin polymerization assay; anti-tumour; cytostatic; ds.	
XX		
OS	Sorangium cellulosum.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	3..992
FT		/tag= a
FT		/label= ORF_A
FT		/product= transposase
FT		/note= "not part of the PKS"
FT	CDS	989..1501
FT		/tag= b
FT		/label= ORF_B
FT		/product= transposase
FT		/note= "not part of the PKS"
FT	CDS	1998..6263
FT		/tag= c
FT		/label= epocA_gene
FT		/note= "encodes the loading domain"
FT	misc_RNA	2031..3548

```
FT FT /tag= d
FT FT /note= "encodes ketide synthase (KS-Y) of the loading
FT FT domain"
FT FT misc_RNA
FT FT 3621..4661
FT FT /tag= e
FT FT /note= "encodes acyl transferase (AT) of the loading
FT FT domain"
FT FT misc_RNA
FT FT 4917..5810
FT FT /tag= f
FT FT /note= "encodes enoyl reductase (ER) of the loading
FT FT domain, potentially involved in formation of the
FT FT thiazole moiety"
FT FT misc_RNA
FT FT 5856..6155
FT FT /tag= g
FT FT /note= "encodes acyl carrier protein (ACP) of the loading
FT FT domain"
FT FT CDS
FT FT 6260..10493
FT FT /tag= h
FT FT /label= epob_gene
FT FT /note= "encodes module 1, the NRPS module"
FT FT misc_RNA
FT FT 2031..3548
FT FT /tag= i
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 2031..3548
FT FT /tag= j
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 6861..6887
FT FT /tag= k
FT FT /note= "encodes heterocyclization signature sequence"
FT FT misc_RNA
FT FT 6861..6887
FT FT /tag= l
FT FT /note= "encodes condensation domain C4 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 7358..7366
FT FT /tag= m
FT FT /note= "encodes condensation domain C7 (partial) of the
FT FT NRPS module"
FT FT misc_RNA
FT FT 7898..7921
FT FT /tag= n
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT misc_RNA
FT FT 7898..7921
FT FT /tag= o
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT misc_RNA
FT FT 8261..8308
FT FT /tag= p
FT FT /note= "encodes adenylation domain A3 of the NRPS module"
FT FT misc_RNA
FT FT 8411..8422
FT FT /tag= q
FT FT /note= "encodes adenylation domain A4 of the NRPS module"
FT FT misc_RNA
FT FT 8861..8905
FT FT /tag= r
FT FT /note= "encodes adenylation domain A6 of the NRPS module"
FT FT misc_RNA
FT FT 8966..8983
FT FT /tag= s
FT FT /note= "encodes adenylation domain A7 of the NRPS module"
FT FT misc_RNA
FT FT 9090..9179
FT FT /tag= t
FT FT /note= "encodes adenylation domain A8 of the NRPS module"
FT FT misc_RNA
FT FT 9183..9992
FT FT /tag= u
FT FT /note= "encodes oxidation region for forming thiazole"
FT FT misc_RNA
FT FT 10121..10138
FT FT /tag= v
FT FT /note= "encodes adenylation domain A10 of the NRPS
FT FT module"
FT FT misc_RNA
FT FT 10261..10306
FT FT /tag= w
FT FT /note= "encodes thiolation domain (PCP) of the NRPS
FT FT module"
FT FT CDS
FT FT 10639..16137
FT FT /tag= x
```

```
FT FT /label= epoc_gene
FT FT /note= "encodes module 2"
FT FT misc_RNA
FT FT 10654..12033
FT FT /tag= y
FT FT /note= "encodes KS2, the KS domain of module 2"
FT FT misc_RNA
FT FT 12250..13287
FT FT /tag= z
FT FT /note= "encodes AT2, the AT domain of module 2"
FT FT misc_RNA
FT FT 13327..13899
FT FT /tag= aa
FT FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT FT module 2"
FT FT misc_RNA
FT FT 14962..15756
FT FT /tag= ab
FT FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT FT module 2"
FT FT misc_RNA
FT FT 15763..16008
FT FT /tag= ac
FT FT /note= "encodes ACP2, the ACP domain of module 2"
FT FT CDS
FT FT 16134..37907
FT FT /tag= ad
FT FT /label= epod_gene
FT FT /note= "encodes modules 3-6"
FT FT misc_RNA
FT FT 16425..17606
FT FT /tag= ae
FT FT /note= "encodes KS3"
FT FT misc_RNA
FT FT 17817..18857
FT FT /tag= af
FT FT /note= "encodes AT3"
FT FT misc_RNA
FT FT 19581..20396
FT FT /tag= ag
FT FT /note= "encodes KR3"
FT FT misc_RNA
FT FT 20424..20642
FT FT /tag= ah
FT FT /note= "encodes ACP3"
FT FT misc_RNA
FT FT 20706..22082
FT FT /tag= ai
FT FT /note= "encodes KS4"
FT FT misc_RNA
FT FT 22296..23336
FT FT /tag= aj
FT FT /note= "encodes AT4"
FT FT misc_RNA
FT FT 24069..24647
FT FT /tag= ak
FT FT /note= "encodes KR4"
FT FT misc_RNA
FT FT 24867..25151
FT FT /tag= al
FT FT /note= "encodes ACP4"
FT FT misc_RNA
FT FT 25203..26576
FT FT /tag= am
FT FT /note= "encodes KS5"
FT FT misc_RNA
FT FT 26793..27883
FT FT /tag= an
FT FT /note= "encodes AT5"
FT FT misc_RNA
FT FT 27966..28574
FT FT /tag= ao
FT FT /note= "encodes DH5"
FT FT misc_RNA
FT FT 29433..30287
FT FT /tag= ap
FT FT /note= "encodes ER5"
FT FT misc_RNA
FT FT 30321..30869
FT FT /tag= aq
FT FT /note= "encodes KR5"
FT FT misc_RNA
FT FT 31077..31373
FT FT /tag= ar
FT FT /note= "encodes ACP5"
FT FT misc_RNA
FT FT 31440..32807
FT FT /tag= as
FT FT /note= "encodes KS6"
FT FT misc_RNA
FT FT 33018..34067
FT FT /tag= at
FT FT /note= "encodes AT6"
FT FT misc_RNA
FT FT 34107..34676
FT FT /tag= au
```

```
FT /note= "encodes DH6"
FT misc_RNA
FT 35760..36641
FT /*tag= av
FT /note= "encodes ER6"
FT misc_RNA
FT 36705..37256
FT /*tag= aw
FT /note= "encodes KR6"
FT misc_RNA
FT 37470..37769
FT /*tag= ax
FT /note= "encodes ACP6"
FT CDS
FT 37912..49308
FT /*tag= ay
FT /label= epoE_gene
FT /note= "encodes modules 7 and 8"
FT misc_RNA
FT 38014..39375
FT /*tag= az
FT /note= "encodes KS7"
FT misc_RNA
FT 39589..40626
FT /*tag= ba
FT /note= "encodes AT7"
FT misc_RNA
FT 41341..41922
FT /*tag= db
FT /note= "encodes KR7"
FT misc_RNA
FT 42181..42423

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No.6.3e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTTCGGGATGATACGCCGCGGTGCGAGTGCATGCGACGAGCGACGACGAGC 60
Db 34350 GCCTTCGGGATGATACGCCGCGGTGCGAGTGCATGCGACGAGCGACGACGAGC 34409
QY 61 GCCTTCGCAATTCACGTTGCCAGCCGGGTGCCGGGCCACGG 101
Db 34410 GCCTTCGCAATTCACGTTGCCAGCCGGGTGCCGGGCCACGG 34450

RESULT 2
AA255887
ID AA255887 standard; DNA; 68750 BP.
XX
AC AA255887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key
FT CDS
FT 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS
FT /*tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT complement (5612..5592)
FT /*tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /*tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /*tag= f
```

```
FT /product= "Type I polyketide synthase, EPOS A
FT (AAV58573)"
FT 11872..116104
FT /*tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P
FT (AAV58574)"
FT 16251..21749
FT /*tag= h
FT /product= "Type I polyketide synthase, EPOS B
FT (AAV58575)"
FT 21746..43519
FT /*tag= i
FT /product= "Type I polyketide synthase, EPOS C
FT (AAV58576)"
FT 43524..54920
FT /*tag= j
FT /product= "Type I polyketide synthase, EPOS D
FT (AAV58577)"
FT 54935..62254
FT /*tag= k
FT /product= "Type I polyketide synthase, EPOS E
FT (AAV58578)"
FT 62369..63628
FT /*tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT (AAV58579)"
FT 63779..64333
FT /*tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT complement (63853..64290)
FT /*tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT 64363..64920
FT /*tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT complement (64287..64727)
FT /*tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT 65063..65767
FT /*tag= q
FT /product= "Orf 10 protein"
FT complement (65008..65874)
FT /*tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT complement (65871..66338)
FT /*tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT 66667..67137
FT /*tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT 67334..68251
FT /*tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT 68346..68750
FT /partial
FT /*tag= v
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
FT W09966028-A2.
XX
PD 23-DEC-1999.
XX
PF 16-JUN-1999; 99WO-EP04171.
XX
PR 18-JUN-1998; 98US-0099504.
PR 24-SEP-1998; 98US-0101631.
PR 05-FEB-1999; 99US-0118906.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

```
XX WPI: 2000-097741/08.
DR P-PDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.
XX
XX New isolated epoethione synthase genes, used for the recombinant
PT production of epoethione for use in cancer therapy -
XX
XX Claim 14: Page 87-104, 174pp: English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epoethione macrolactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epoethiones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epoethiones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epoethiones as anticancer agents, they are problematical to produce on a
CC large scale. Epoethiones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epoethiones. The nucleic acids of the invention may be
CC used for the recombinant production of epoethiones in a heterologous host
CC that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
SQ
XX
XX Query Match 98.4%; Score 99.4; DB 21; Length 68750;
XX Best local Similarity 99.0%; Pred. No. 1.8e-19;
XX Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGACAGTATGGCGACCGAGCAGCAGGC 60
DB 39962 GCCTTGGCGGATGATACGCCGGGCGGTGACAGTATGGCGACCGAGCAGCAGGC 40021
XX
XX 61 CGCCTGCAATTCCACGTTGGAGCCGGGTGCCGGCCACGG 101
DB 40022 CGCCTGCAATTCCACGTTGGAGCCGGGTGCCGGCCACGG 40062
XX
XX RESULT 3
XX ID AAI99683 standard; DNA: 4403765 BP.
XX AC AAI99683;
XX
XX 15-JAN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
XX Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX US6294328-B1.
XX PN 25-SEP-2001.
XX PD 24-JUN-1998; 98US-0103840.
XX PF
```

```
XX
XX 24-JUN-1998; 98US-0103840.
PR
XX
XX (GENO-) INST GENOMIC RES.
PA
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
PI
XX WPI: 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 4: SEQ ID NO 2; 3pp + Sequence Listing: English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.
XX
XX Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
SQ
XX
XX Query Match 31.3%; Score 31.6; DB 22; Length 4403765;
XX Best local Similarity 58.5%; Pred. No. 8.9;
XX Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
XX
XX 1 GCCTTGGCGGATGATACGCCGGTGGCGGTGACAGTATGGCGACCGAGCAGCAGGC 60
DB 4263192 GCCTTGGCGGATGATGACGCCGGTGGCGGATGACAGCAGCAGCAGCAGC 4263133
XX
XX 61 CGCCTGCAATTCCACGTTGGAGCCGGGTGCCGG 94
DB 4263132 GGCCTGCAATTCCACGTTGGAGCCGGGTGCCGG 4263099
XX
XX RESULT 4
XX ID AAI99682 standard; DNA: 4411529 BP.
XX AC AAI99682;
XX
XX 15-JAN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX US6294328-B1.
XX PN 25-SEP-2001.
XX PD 24-JUN-1998; 98US-0103840.
XX PR 24-JUN-1998; 98US-0103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX PF WPI: 2001-647261/74.
XX
```

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
PS
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
CC
XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
SQ
Query Match 31.3%; Score 31.6; DB 22; Length 4411529;
Best Local Similarity 58.5%; Pred. No. 8.9;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 GCCTTCGGGATGATACGCCGGTGGCGGTGACATGCGAGCGAGCGACCGAGC 60
DB 4270935 GCCCTGCTGGAAATATCTGACGTGCGCAATCCAGAACAGCCATCTCTCGCC 4270876
QY 61 CGCCTGCAATTCACGTTGCGGCGGCGGCGCGG 94
DB 4270875 GGGCCCGAGCACATCTTTTCGATCTCGAATCGG 4270842
RESULT 5
ABLO6373
ID ABLO6373 standard; cDNA: 947 BP.
XX
AC ABLO6373;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13601.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PD 23-MAR-2001; 2001WO-US09231.
XX
PE 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB62270.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 13601; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLO1840-ABLO30511), expressed DNA
CC sequences (ABLO1840-ABLO16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 947 BP; 232 A; 265 C; 291 G; 159 T; 0 other;
SQ
Query Match 30.1%; Score 30.4; DB 23; Length 947;
Best Local Similarity 61.2%; Pred. No. 9.8;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 7 GCGGATGATACGCCGGTGGCGGTGACATGCGAGCGAGCGAGCGCGCTG 66
DB 357 GCCGAGGAAACTGCGCGAGGTGCGATGCGAGCTGATGAGAGAGCCCGGACAGGCTG 416
QY 67 CAATTCACGTTGCGAGCCG 86
DB 417 CACGTCTCGATGTGGCGAG 436
RESULT 6
ABLO6348
ID ABLO6348 standard; cDNA: 2890 BP.
XX
AC ABLO6348;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13526.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PD 23-MAR-2001; 2001WO-US09231.
XX
PE 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB62245.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 13526; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLO1840-ABLO30511), expressed DNA
CC sequences (ABLO1840-ABLO16175) and the encoded proteins
CC (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 2890 BP; 784 A; 688 C; 704 G; 714 T; 0 other;

Query Match 30.1%; Score 30.4; DB 23; Length 2890;
Best Local Similarity 61.2%; Pred. No. 11;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 7 GCGGATGATACGCCGCTGCGGTGATGCGACGAGGAGCGCCGCTG 66

DB 2484 GCCGAGGAAACTGGGCGAGTGCAGCTGATGGAGAGACCGCCGACAGGCTG 2543

QY 67 CAATTCACGCTTGCAGCCG 86

DB 2544 CACGTCCTGGATGTGGCGAG 2563

RESULT 7
ABLO6372
ID ABL06372 standard; cDNA; 2947 BP.

XX ABL06372;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13598.

KM Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB62269.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 13598; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABLO1840-ABL16175) and the encoded proteins

CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2947 BP; 679 A; 726 C; 799 G; 743 T; 0 other;

Query Match 30.1%; Score 30.4; DB 23; Length 2947;

Best Local Similarity 61.2%; Pred. No. 11;

Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 7 GCGGATGATACGCCGCTGCGGTGATGCGACGAGGAGCGCCGCTG 66

DB 1357 GCCGAGGAAACTGGGCGAGTGCAGCTGATGGAGAGACCGCCGACAGGCTG 1416

QY 67 CAATTCACGCTTGCAGCCG 86

DB 1417 CACGTCCTGGATGTGGCGAG 1436

RESULT 8
ABLO8162
ID ABL08162 standard; cDNA; 4282 BP.

XX ABL08162;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18968.

KM Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB64059.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 18968; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABLO1840-ABL16175) and the encoded proteins

CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4282 BP; 966 A; 1112 C; 1156 G; 1048 T; 0 other;

Query Match 30.1%; Score 30.4; DB 23; Length 4282;

Best Local Similarity 61.2%; Pred. No. 11;

Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 7 GCGGATGATACGCCGCTGCGGTGATGCGACGAGGAGCGCCGCTG 66

DB 270 GCCGAGGAAACTGGGCGAGTGCAGCTGATGGAGAGACCGCCGACAGGCTG 329

QY 67 CAATTCACGCTTGCAGCCG 86

DB 330 CACGTCCTGGATGTGGCGAG 349


```

RESULT 9
ABL64403
ID ABL64403 standard; DNA: 167343 BP.
XX
XX ABL64403;
XX
DT 15-MAY-2002 (first entry)
XX
DE Stomach cancer related gene sequence SEQ ID NO:2740.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200194629-A2.
PN
XX 13-DEC-2001.
PD
XX
XX 30-MAY-2001; 2001WO-US10838.
PF
XX
XX 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.

```

```

XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 2740; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
XX Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 13 other;
SQ
XX
XX Query Match 29.7%; Score 30; DB 24; Length 167343;
XX Best Local Similarity 67.7%; Pred. No. 20;
XX Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
XX
XX
XX 40 GCGACCGGAGGAGCGACCGCGCTGCATTCACGTTGGAGCGCGGTCGCCGCCAC 99
QY ||| ||||| ||| ||||| ||||| ||| ||| ||| ||||| |||||
Db 117079 GCGTCCGAAGACCGCGGAGCTCCGCGAGCTCCGCGGTACACCGGGTACGGGCGG 117138
QY 100 GG 101
Db 117139 GG 117140
XX
XX
XX RESULT 10
XX ABL67239
XX
XX ABL67239 standard; DNA: 167343 BP.
XX
XX
XX 15-MAY-2002 (first entry)
XX
XX Thyroid cancer related gene sequence SEQ ID NO:5576.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KM cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX WO200194629-A2.
PN
XX 13-DEC-2001.
PD
XX
XX 30-MAY-2001; 2001WO-US10838.
PF
XX
XX 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.

```

PR 25-SEP-2000; 2000US-23492AP.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;

WPI: 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -

Claim 1; SEQ ID 5576; 44pp; English.

CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytoskeletal
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.

SQ Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 13 other;

Query Match 29.7%; Score 30; DB 24; Length 167343;

Best Local Similarity 67.7%; Pred. No. 20;

Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 40 CGCAGCAGAGGAGCAGCCGCCCTGCAATTCAGCTTCGCGCGGCGGCGGCAC 99

DB 117079 GCGTCCGAGAGCGCGAGCCTCCGCGAGCTCCGCGTACACCGCGGTGAGGCGCG 117138

QY 100 GG 101
 II
 DB 117139 GG 117140

RESULT 11
 AA253393/c
 ID AA253393 standard; DNA; 1620 BP.

AC AA253393;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 212 partial DNA sequence SEQ ID NO:735.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.

OS Neisseria meningitidis.

PN MO9957280-A2.

PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.

PI (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizzo M, Rappelli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

PI P-PSDB; AA74631.

DR WPI: 2000-062150/05.

DR Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

PS Claim 7; Page 488-489; 1453pp; English.

CC AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254773 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC immunization of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

SQ Sequence 1620 BP; 412 A; 625 C; 318 G; 260 T; 5 other;

Query Match 29.1%; Score 29.4; DB 21; Length 1620;

Best Local Similarity 53.7%; Pred. No. 20;

Matches 51; Conservative 4; Mismatches 40; Indels 0; Gaps 0;

QY 6 CCGGATGATACCGCGGTGCGGTGAGTATGCGCAGCAGGAGCAGCAGCGCGCT 65

Db 1588 CGCGATGATGTCGCGGTTGGGCTGTAGCGCGCGAGCGCTGTAAAGCGATGCG 1529
OY 66 GCAATTCACGCTGGAGCGCGGCGCGGCGACG 100
Db 1528 GCARCGCTASGATTTGGGCTGCAMTGCAGCGCGG 1494

RESULT 12

AAV58938
ID AAV58938 standard; DNA; 10095 BP.

XX AAV58938;

XX 15-FEB-1999 (first entry)

XX Mycobacterium tuberculosis embCAB operon.

XX Drug resistance; antibiotic resistance; antimycobacterial;

KW ethambutol; embCAB operon; infection; vaccine; therapy;

KW tuberculosis; ds.

OS Mycobacterium tuberculosis.

XX WO9841533-A1.

XX 24-SEP-1998.

XX 16-MAR-1998; 98WO-US05128.

XX 20-MAR-1997; 97US-0822586.

XX (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

XX Jacobs WR, Musser J, Telenti A;

XX WPI; 1998-521160/44.

XX P-PSDB; AAM73052-54.

XX Wild type and mutated sequences of Mycobacterium embCAB operon -
PT useful to e.g. identify ethambutol-resistant mycobacterial strains
PS and produce antisense sequences to treat mycobacterial infections
XX Disclosure; Fig 4A-I; 62pp; English.

XX This is the DNA sequence of the Mycobacterium tuberculosis embCAB
CC operon, which determines resistance to the antimycobacterial drug
CC ethambutol (EMB). It includes the embA, embC and embB genes
CC that encode proteins (see AAM73052-54) which are the target of
CC action of M. tuberculosis for EMB. Wild-type and mutated embCAB
CC nucleic acid sequences are useful e.g. as probes used in the
CC diagnosis of drug-resistant mycobacteria or to determine the
CC susceptibility of mycobacteria to EMB. The nucleic acids are also
CC useful in the treatment of mycobacterial infections; anti-DNA or
CC anti-RNA sequences can be administered to inhibit embCAB operon
CC mRNA activity (claim). The invention additionally provides for
CC the use of embCAB operon nucleic acid sequences as vaccines, or
CC to improve existing vaccines.

XX Sequence 10095 BP; 1469 A; 3475 C; 3370 G; 1780 T; 1 other;

XX Query Match 28.7%; Score 29; DB 19; Length 10095;

XX Best Local Similarity 55.4%; Pred. No. 30;

XX Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGAGTCAATGCGACGAGACGACGACG 60

OY 61 GCGCTGCAATTCACGTTGCGAGCGCGGTCGCGGCGACG 101

OY 925 CACGCGGCTTCGCGCGCGCTGCTGTCGACGCGCGG 965

RESULT 13
ID AAF88338
XX AAF88338 standard; DNA; 14775 BP.

XX AAF88338;

XX 28-AUG-2001 (first entry)

XX S. spinosa DNA fragment encoding ORF21, SEQ ID 47.

KW Forosamine; trimethylrharnose; polyketide synthase; biosynthesis;

KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;

KW macrolide; insecticidal; polyketide synthase; ds.

XX Saccharopolyspora spinosa.

XX DE19957268-A1.

XX 08-MAR-2001.

XX 29-NOV-1999; 99DE-1057268.

XX 27-AUG-1999; 99DE-1040596.

XX (FARB) BAYER AG.

XX Eberz G, Moehrl V, Froede R, Velten R, Salas JA;

XX WPI; 2001-267102/28.

XX P-PSDB; AAB70968.

XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

XX recombinant production of insecticidal spinosyns and their derivatives

XX Claim 7a; Page 239-264; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrharnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC also used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 21, encodes an S. spinosa polyketide synthase.

XX Sequence 14775 BP; 1969 A; 4181 C; 5731 G; 2894 T; 0 other;

XX Query Match 28.5%; Score 28.8; DB 22; Length 14775;

XX Best Local Similarity 58.0%; Pred. No. 36;

XX Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 6 CGCGATGATGCGCGGCGGTGCGAGTCAATGCGCGAGGAGGAGCGCGGCT 65

OY 4095 CGTGACGCGCACCTGAGTGTGCGGCGCGCGCGCGCGGCGGAGGAGAAC 4154

OY 66 GCAATTCACGCTGGAGCGCGGTCGCG 93

OY 4155 GCAGTGGCACTTCGCGCGCGCGTGGCG 4182

RESULT 14

AAF88313
ID AAF88313 standard; DNA; 50000 BP.

XX	AAF88313;
AC	
XX	
DT	28-AUG-2001 (first entry)
XX	
DE	S. spinosa DNA fragment SEQ ID 2.
XX	
KW	Forosamine; trimethylrhinose; polyketide synthase; biosynthesis;
KV	spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW	macrolide; insecticidal; ds.
XX	
OS	Saccharopolyspora spinosa.
PN	DE19957268-A1.
XX	
PD	08-MAR-2001.
XX	
PF	29-NOV-1999; 99DE-1057268.
XX	
PR	27-AUG-1999; 99DE-1040596.
XX	
PA	(FARB) BAYER AG.
XX	
PI	Eberz G, Moehrl V, Froede R, Velten R, Salas JA;
XX	
DR	WPI, 2001-267102/28.
XX	
PT	New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT	recombinant production of insecticidal spinosyns and their derivatives
XX	-
PS	Claim 7; Page 31-49; 354pp; German.
XX	
CC	This invention describes a novel method nucleic acid (I) and its encoded
CC	polypeptide (II) containing at least one region that encodes an enzymatic
CC	activity involved in biosynthesis of spinosyns. (I) are used (1) to
CC	identify, inactivate or modulate genes involved in the biosynthesis of
CC	(II); (11) to generate a library of polypeptide synthases; (111) for
CC	adding forosamine or trimethylrhinose to a spinosyn or polyketide
CC	aglycone; and (1v) for recombinant production of the corresponding
CC	enzymes, which are used for production of (II), their precursors or
CC	derivatives, including production of transgenic plants that express (II)
CC	and thus have increased resistance to insects. (I) are also useful as
CC	markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC	macrolides with insecticidal, but not antibacterial, activity, and can
CC	also be used to raise specific antibodies, useful for identifying
CC	expression clones in a gene bank. Cells transformed with (I) may produce
CC	(II) at significantly increased levels or produce new derivatives of
CC	(II). This sequence represents a genomic DNA fragment of the S. spinosa
CC	genome which contains the coding regions for proteins involved in
CC	forosamine, trimethylrhinose and polyketide synthase biosynthesis.
XX	
SO	Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 other;
Query Match	28.5%; Score 28.8; DB 22; Length 50000;
Best Local Similarity	58.0%; Pred. No. 40;
Matches	51; Conservative 0; Mismatches 37; Indels 0; Gaps 0
OY	6 CCCGATGATACCGCGGTGGCGGTGCAGTACTGCCAGCAGAGACGACCGCCCT 65
DB	23701 CCTGACGCGCACACCTGAGTCCTGTGCGGCGTTGGCGGCGCGGTGAGAGAAC 23760
OY	66 GCAATTCCAGCTTGCGAGCGCGGTGGCGG 93
DB	23761 GCAGCTGCACATTGCGCGCGGCGGTGGC 23788
RESULT 15	
ID	AAF88316
XX	AAF88316 standard; DNA; 50000 BP.
NC	
XX	AAF88316;

```

DE      S. spinosa DNA fragment SEQ ID 5.
XX
XX      28-AUG-2001 (first entry)
DT
XX
KW      Forsamine; trimethylrhannose; polyketide synthase; biosynthesis;
RV      spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX      macrolide; insecticidal; ds.
OS      Saccharopolyspora spinosa.
PM      DE19957268-Al.
PD      08-MAR-2001.
PF      29-NOV-1999;   99DE-1057268.
PR      27-AUG-1999;   99DE-1040596.
XX      (FARB ) BAYER AG.
PI      Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
DR      WFI: 2001-267102/28.
XX
PT      New nucleic acid encoding enzymes for spinsyn biosynthesis, useful for
PT      recombinant production of insecticidal spinsyns and their derivatives
PS      -
XX
XX      Claim 7; Page 74-91; 35app; German.
XX
XX      This invention describes a novel method nucleic acid (I) and its encoded
CC      polypeptide (II) containing at least one region that encodes an enzymatic
CC      activity involved in biosynthesis of spinsyns. (I) are used (i) to
CC      identify, inactive or modulate genes involved in the biosynthesis of
CC      (ii); (iii) to generate a library of polyketide synthases; (iiii) for
CC      adding forsamine or trimethylrhannose to a spinsyn or polyketide
CC      aglycone; and (iv) for recombinant production of the corresponding
CC      enzymes, which are used for production of (II), their precursors or
CC      derivatives, including production of transgenic plants that express (II)
CC      and thus have increased resistance to insects. (I) are also useful as
CC      markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC      macrolides with insecticidal, but not antibacterial, activity, and can
CC      also be used to raise specific antibodies, useful for identifying
CC      expression clones in a gene bank. Cells transformed with (i) may produce
CC      (ii) at significantly increased levels or produce new derivatives of
CC      (ii). This sequence represents a genomic DNA fragment of the S. spinosa
CC      genome which contains the coding regions for proteins involved in
CC      forsamine and trimethylrhannose biosynthesis.
XX
XX
SQ      Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 other;
Query Match          28.5%; Score 28.8; DB 22; Length 50000;
Best Local Similarity    58.0%; Pred. No. 40;
Matches     51; Conservative    0; Mismatches     37; Indels       0; Gaps        0
OY           6 CGCGATGATAACGCCGGTGTGCAGGTTCATTGGACCGAGCAGCACCGCTT 65
            ||||| | | | | | |||| | | ||| | |||| | 
DB 28077 CGTGAGCGCAACTTAAGTCGTGGTAGGCGCTTCCC GCCCGCTGCACAGAACAACC 28136
            || | | | | | | | | | | | | | | | 

OY         66 GCAA TTCACAGTTGCGA GCCGGGTGCGC 93
            || | | | | | | | | | | | | | | | | 
DB 28137 GCAGCTGGCACTTCGGCGCGCGCTGGCGC 28164


Search completed: November 6, 2002, 12:46:48
Job time : 1760 secs
```

Search completed: November 6, 2002, 12:46:48
Job time : 1760 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_34350_34450

Sequence: 1 gcttcgcgatgataccg.....agccgggtgcgggcccacg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl:*
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11:  gb_sts:*
12:  gb_sy:*
13:  gb_un:*
14:  gb_vi:*
15:  em_ba:*
16:  em_fun:*
17:  em_hum:*
18:  em_in:*
19:  em_mu:*
20:  em_om:*
21:  em_or:*
22:  em_ov:*
23:  em_pat:*
24:  em_ph:*
25:  em_pl:*
26:  em_ro:*
27:  em_sts:*
28:  em_un:*
29:  em_vi:*
30:  em_htg_hum:*
31:  em_htg_inv:*
32:  em_htg_other:*
33:  em_htg_mus:*
34:  em_htg_pln:*
35:  em_htg_rod:*
36:  em_htg_mam:*
37:  em_htg_vrt:*
38:  em_sy:*
39:  em_htgo_hum:*
40:  em_htgo_mus:*
41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	71989	6 AR172664	AR172664 Sequence
3	99.4	98.4	68730	1 AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6 AR193029	AR193029 Sequence
5	99.4	98.4	68750	6 AR199551	AR199551 Sequence
6	99.4	98.4	68750	6 AR199559	AR199559 Sequence
7	99.4	98.4	68750	6 AR199567	AR199567 Sequence
8	99.4	98.4	68750	6 AR201097	AR201097 Sequence
9	99.4	98.4	68750	6 AR208671	AR208671 Sequence
10	65.2	64.6	16124	1 AX024384	AX024384 Sequence
11	65.2	64.6	16124	6 AX024277	AX024277 Sequence
12	34.4	34.1	609	4 AY112896	AY112896 Equus cab
13	31.8	31.5	10507	1 AE012031	AE012031 Xanthomon
14	31.6	31.3	17205	1 AE007185	AE007185 Mycobacte
15	31.6	31.3	23740	1 MTW026	MTW026 Mycobacte
16	31.6	31.3	96774	2 AP004163	AP004163 Oryza sat
17	31.6	31.3	148766	2 AC118895	AC118895 Rattus no
18	31.4	31.4	37445	1 SC8A6	SC8A6 Streptomy
19	31.2	30.9	11910	1 AE012015	AE012015 Xanthomon
20	31.2	30.9	310050	1 RME603642	RME603642 Rhizobium
21	31	30.7	409	14 AF065707	AF065707 Stealth v
22	31	30.7	845	14 AF065691	AF065691 Stealth v
23	31	30.7	215050	2 AL646057	AL646057 Ralstonia
24	30.8	30.5	235571	2 AC128364	AC128364 Rattus no
25	30.6	30.3	2947	1 SLI18818	SLI18818 Streptomyce
26	30.6	30.3	3012	10 AF293383	AF293383 Rattus no
27	30.6	30.3	4851	1 SC018817	SC018817 Streptomyce
28	30.6	30.3	12647	1 AE012088	AE012088 Xanthomon
29	30.6	30.3	29080	1 SCBAC2861	AL593842 Streptomy
30	30.6	30.3	40337	2 SC5A7	AL031107 Streptomy
31	30.6	30.3	140714	1 RN374E16	AL603726 Rattus no
32	30.6	30.3	190050	2 AL646059	AL646059 Ralstonia
33	30.4	30.1	1078	3 AY051416	AY051416 Drosophila
34	30.4	30.1	106186	8 AF448416	AF448416 Zea mays
35	30.4	30.1	123751	2 AC019937	AC019937 Drosophila
36	30.4	30.1	153749	8 AP003019	AP003019 Oryza sat
37	30.4	30.1	156508	3 AC007581	AC007581 Drosophila
38	30.4	30.1	171151	3 AC007574	AC007574 Drosophila
39	30.4	30.1	226001	8 AF391808	AF391808 Zea mays
40	30.4	30.1	310958	3 AE003464	AE003464 Drosophila
41	30.2	29.9	11169	1 AE004741	AE004741 Pseudomon
42	30.2	29.9	84487	2 OSJN00006	AL606451 Oryza sat
43	30.2	29.9	115548	2 OSJN00010	AL606453 Oryza sat
44	30	29.7	2137	9 AB060236	AB060236 Macaca fa
45	30	29.7	2212	1 BS068411	BS068411 Burkholderia

ALIGNMENTS

```
RESULT 1
AF217189
LOCUS
DEFINITION
AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds;
putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189
KEYWORDS
AF217189.1 GI:7453554
SOURCE
ORGANISM
Polysangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangium; Polysangium; Polysangiaceae; Polysangium.
REFERENCE
1 (bases 1 to 58733)
```

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epochlone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649993
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epochlone biosynthetic gene
cluster from *Sorangium cellulosum*
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20283058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submision
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1. 58733
/organism="Polyanthum cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1. 992
/note="orfa"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:745355"
/translation="VRGHEAFADFGVPRVLLYDNLKNAVERHGDAIRFHTLLAT
SADYFEPFPPVAVARGNEKGRVERAIRVREGFPERARADLGDINRQATWNTSSAL
DTSWEDRARVQAQFDDERSVLRHPOTPEPDHEVEVEGTCPYARPDLDVYSPH
DSTRRLVYADLSQVRADGNQIVATVHYRSMDDQIEOPHILORLVDEKRRARERH
GIDRLARARASQAFRLIYAERGDNVGSIALRLQILDVGAELDEAVYELERDTI
HIGAVROYIDRRSRERHLPPIVIPYTRGEHALVYTPHSLTYDLKADTP"
989. 1501
/note="orfb"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:745356"
/translation="MTDLPTETKDKLSLGLGLACMEQLADKPMLEREVLAIERE
RHRSLERLRKNSRYAAAFKPMTPDPSMPKIDREAVDOLYDSRYADLFEVYTRVD
AOKPLLSTNRKAFADMGQVYFAACVTLVDRLVYRAVEITEASRYLKEAKELNAT
RTKORRTKKH"
1998. 6263
/gene="epoa"
1998. 6263
/gene="epoa"
/note="type I polyketide synthase; epochlone PKS loading
module"
/codon_start=1
/transl_table=11
/product="Epoa"
/protein_id="AAF62880.1"
/db_xref="GI:745357"
/translation="MADRIERAAEDPIAIYAGCRLPGVITDSGFWTLLGSRDVT
GOVPERMDAAWFPDLDAPCKTPTVTRASFISDYACPDASFSGISPRALMDGPHR
LLEVCWELENAALAPSAVLGTETGVLGIGIPQCEALPAPATASAEIDAGLGTM
PSVGAGRISYVGLRGPCVAVDTASSSLVAVHALCOSLRSECGSTALAGVSLMSP
STWLVKTRIALATDGRCAFAEADGFCRGCACVAVYKRLSGARADDDRIIAVYRG
SAIINDGASGGLTVNGSIOETVLKRALADAGCAASVYGAHGTTIGDPIEIOA
LNAVYGLRDVATPLLIGSKVTNLGHPYASGITGLKVVLSLQHQIIPAHHAQALN
PRISGDLTLVTRARTPMDNTRPGKVSFGMSGTAHVHLEAPATCTPPAPE
RPAELLVLSARTAAALDAHAARLDHLETPSQCLGVAFSLATRSAMERHAAVAT
SSEGRALDAAGCOTPGVVYRGIAIDSSRGRLAFLFTGOGAQTGLMGKGLDVMPAF
RAFDLCVFLFNOELDRPLREYMAEPASVDALADDOFAOTPALFTFETALALMRS
WGVEPELVAGHSIGELVACVAGVPSLEDAVFLVAARGLMALPAGGMAVSTAAPEA
DYAAAVAPHAASVSTAAVNGPQVVIYAGAGQVYHIAAAMARGARTRKLIHSHAFS
PLMAPLEHFGRAVSAVSTRPSIVLSMLSGAKGTDESSPGYWRVRIAREVVRADG
VKALHAAGAGTVEVGPKSTLLGLVPACLPDARPALLLASSRAGRDPEALGLGLW

AVGLVSNAGLEPSSGRRVLPPTYPMORERYWIDTKADDAAGDRRAGAGHDEVKQ
GAVRGDRRSARLIDHPPESSGRRKVEAAGDPFLLEIDEVDRLRLVYTERNAPG
LGEVELVADAAGLSENDVOLATGAMPDLPKGNPNPLLEGGCAGCACTIYAGVYAGV
VGOPYTALSAGAFATHYTSALVYLPQALISATPAAMPVAYLTAAVADGSIAPLOP
GERVLIHAATGCGVGLAAYOMAOHVCAEVAHATRGTEPKRATLESLEYVDSRSRPV
ADVRAWTGGEVDVYVNLISGELIDKSEFLLSHORFVELGRDVCYADNOQLGFRFLR
NLSFSLVDLRGMMLEPRPARVRALEFELLGLIAAGVTPPIATLPIIAVVALEFRMAQ
AOHLGKVLTLIGDPEVOIRIPYHAGAGSTGDRDLDRLASAAPARAALAEFRATQ
VSOVLRTPETIKGAELFTRLGMSLMAVETSRRLTEASIKLKSTTFLSTPNIALLT
ONLDAALFALSLEVAENLHAGVQSDVSSGADQWDEIIAL"
6260. 10492
/gene="epob"
6260. 10492
/gene="epob"
/note="non ribosomal peptide synthetase; epochlone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:745358"
/translation="MTINOLNLEHOGYKLAADGERLQIOAPKNAIPLNLIARISEH
KSTITITMLRQRLPAESIYPAPABRRVPPPLDITQGSYTLGRGATTVPSGTHAREYD
CTDLVVARLSRAFRVARHDMLRATHLPDMQVLEPVADIEITDLRGDRSTREA
RLVSLRDAMSHRIYDTERPPLVAVAVRLDEQOTRLVSTIDLINVLGSLSTIFKDWL
SFYEDPETSLPYLEISYRDYVLALESRRKSEHQSMQWYKRRVALEPPMLPKKAD
PSTLEIRFRHTEOMLPDSMSRLKORVGERGLPTGYIALAFSEVIGRMSASPERTL
NITLFNRLPVHPRVNDITGDFTSMWLLIDITRDKSFPQAKRIOELOPEMDADDS
GIEVOREARVYGIORCALFPVVLISALNOQVGYTSQRIGTPTPYTSTORPOLLDH
QLEHGDGLVAMDITDGVFPDLDLMEALVATLKRITTEPWSQKCSLPQAQLE
AASANENNTLSSEHTLHGLFAVAREOLPMOLVARSKTTLYEELSRSRSLGKRL
BOGARPNITVAAYMEKMGEOVAVLALESAAVPIADDTLPAEIRIHLHLGHEVGLV
LTQPLMDGLSWPPIORLEVDACVEGDGDDPMPLOTSPDLAVIYVTGSGSTLEPK
GVIMIDRGAVNTIILINERREIIPGDVATLSLSEFISYDVDFEILAAAGTVVPA
SKTRDPANHAALIEREKYTVMSYALRMVIEHSEGPDSIARSLRSLISGDMPIV
GLPGLQAIRPGCVSISUGATSEASITSGIVKRVNDVSMASIPGRRLNROTFRVLD
EALERPVPWVPGQITGGVGLALGYWROEKRNKSLVHPETEGLYVTGDGRFLPD
GNIEFMGRDNOIKLRGYRVELGEIETELKSHPNRVDVITPVGNDANKLLAVVP
EGTRRRAAEODASLKTERVDARAHAAKADGSDGRVOFKLARHGLRDLGKPAVVD
TGLVREAGLDVYARSRVYRTELEAPIDPEVERGRLSGCSVEPFGALPKRRYPSAG
SPYVOTYAYAKSGRIEYDEGFYHYHPEHRLKVSQHGIERGHAYVONDPYDEAA
FOLLVAGRDAIESLYGSLSRFCILLEGAYNAQLMEBAPSNCINVCYVGQDFEQVR
PVULDRHSOVYVHGMIGGRVDRPQVOTCLGDSSPRAATTGAPPGNDPADLIRD
PLRKLIPYMPTEVEVEDALPLTNSGVYDKRALREKDTSPRSHSGTAPRDALDEI
LVAAVRVSGLEEVGLOQSFVDLGAATSHIYMRSLQKRLDREIATITELQYPMIGS
LASGLRDSKIDQRPNMQDRVEARKGRRS"
10639. 16137
/gene="epoc"
10639. 16137
/gene="epoc"
/note="type I polyketide synthase; epochlone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:745359"
/translation="MEBDESSAIAVIGSGRPPGARDLDEFRNLROGTEAVORESEO
ELASGCPALVLDPNYRAGSVLEDDVRFDAFGCISPREALMOPHRIEMCEAME
ALENNGYDPTAIEGISTGYAGANMSYITLSNLHEPMAKRPGMOTILGNDKIYAT
HVSYRLNLRGPTISVQIACSTSLVAVHLACMSLHDLDECDMAAGITVRIIPRAVYV
AEGGTFSPGHCRAFDANKANGIMONGGVVLLKPLDLASGDDVRAVYVILGSAVND
GARKIGTFAPSEVGOAOAMEALALVAGEARISIOYIETHGCTLLGDAIETALARE
DBDASTRSCAIGSAGISYGIHLESAGIAGLITVYALEHRLPSSLMEESPNSIDF
AASPPYVTSIKDNWTSQPRRAGVSSFGICGTNAHVYLEAPAKLPAAAPARAEAL
FVSKSAALDAALARLDHLOAHQIGSLGVAFSLATTSMPERHAMAAPSREAL
REGIDAAARGPQPPAVRGRCSPGVNVVFPQVSGQWGMGQDLAEAPRFAAL
SACDRAIIEAEMQSLAEIADAESSQIERIDVQVPGQVQAWMGLAELMRSGVGPV
VIGHMGEVAAHVAGALSLEDAVALICRSLRLIRISGOGMAVTELSIAEALR
GYEDVSVAVNSPSTVLSGEPRAIIGTSLSNKKGFCRRVYKDVASHSPQDPLR
EDLLAALGILRRRAAIVPMRSYVTGAMVAGPELGNVYMMNNTROVRAVAYVQAOLOG
GHGLFEVSPHPLITTSYBENKRAAQWGAAGSIRKQODERPALEALGLAMAGTIP
VPMGRLPFAGGRVPLPTYPMQRERYWTEAPAKSAAGDRVRAGGHPDLGEMOTLST

Query Match	Best Local Similarity	100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 1	GCCTTCGGGATGATACGCCGCGTGGCGGTGCGAGTCATGCGGAGCCGAGACGACGACG 60	OTSRIMETITLTKRIPWIGDHRVQGVNVPFGAATYEMALSSGAEALGSDPLOTITDV LAEFLAPAGDAVIVOVYITTEQPSGRLOQFOTIASPAPGASHFVKHARGLLIRVERTE VPALTLISAVRARLQOASMPAAATYAEITTEGLQYGFAPQGIATLMBEGALGVRLP DAGSAEYERLHPALIDLADCFQVYSGVGGGATPNVPYVRLLEDRQSGELTMCAR RVNHHGROTPDROGADFWYVDSGAAVVAEYSGVLAQGLPGVRRREEDDMFLELEMER AAVGTAKVNNAGRIITLGGGGGIGALISMLIAGAHVAAHSAESTSAACVRIALLAKF DGCAPTAVVHIGSLDGGEDPCIGAGOGALIDAPRSADVSPDALIDPALVRDCSVLTV QALAGMFRDAPRIMLITLTKRAQAAVAGADVSVAQPLIGLIRVIAAMEHADRCAVDID PTPRDGTGLALIELLDADAEEVALNAGGRCVARIYRQEPTEPRORIISCVPTDVT IRADSTYLVWGGLGIGLISVAGMLAEAGAHILVYVGRSGAAVEORAAVVALEAGAR VTVAKHVAADRAOLERILREYTTSGMRLRQGVNHAAGLIDDLGIMLPQAPRRRYMAK VOGALHIALTRERAPLSEFVIVYASGVLLSSPQCGANTAAANTEILDALHHRRAQGLPA LSVDMGLFAEYGAAGAAOEDNGARLVYSGMSLTPDEBLSLARSVLESGVQVQGVPNV PRIMVELYPAAGSRMLSRVITAHRAASAGPAGGIDRLIRIAAEPASVGLTEPLLR AOISOVLRYPEAGKIEVDAPLTSIGMSLMQLELNRLEAMIGTIVPATLMTPTVAA LSGLHAREACEAARVESPHTTADSAVEIEEMSDDLTQLIAAKFALT"
Db 34350	GCTTCGCGGATGATACGCCGCGTGGCGGTGCGAGTCATGCGGAGCCGAGACGACG 34409	16134. .37907 /gene="epod" /note="type 1 polyketide synthase; epothilone PKs modules 3, 4, 5, and 6" /codon_start=1 /transl_table=11 /product="epod" /protein_id="AA62883.1" /db_xref="GI:7453560" /translation="MTTRRGPTAONPLKQAAITIORLEERLAGIAQAEERTPEIATV GICGRFPGADAPAPAEWEILIDAEPRDVAOPIDRRMAIVGVAPEVAVPMACILEPIDC PDAPFCISPREARSLDPQHRULILEVAMEGLIEDGITPRISIDSRITVYFGAFLPADIA RVARLDPREERDAYSATGNMISIAAGLSYLLQGLGLVADACSSIVAIIHACSS LRAESEDALAGGVSTLSDPMMEAAARTQALSPDGCRFPDVSANGFVVEGGGLV LKRIISDAOBDGRIWALIRGSAIINHDSRIGLPAVYLAEVYIRLEASHVAGAV DYVTHGTGTSIGDPIEVEALRATVGPARDGTRCVAGAKTVNIGLEAAAGVAGLIK AAISLTHEERIPIRNINPELRIPRIELSESALAIATEPVPRTDPRRAGYSISGMSST NAHVYLEAPAVELMPAPRSARELIVLSKSEBQALDQARLREHIDMPELGLGV AFSLATYRSAMTHRLAAVYTSREGILALSAVAGQDPAPAAARLASSSRGKLLALFT GOGQOTPGMGKGLCAAMPAPAREAFEDRCVTLFDRILDPLEEVMAEAGSASELLDQT AFQPALFAVEYALITALMRMGVPEELVCHTSIGELVAACVAGVFESLDCVRLVARG RLMOGLSAGGAMVSLGAPAEAVAAVPAHAWYSIAVNPDEQVYIAGVCAVOALIAA GFAPARVTRILHVSNAFHSPLMEPLMEERGVAASATYTRPSVSLVSNLSGKVTYDE
OY 61	CGCGTGCATTCACAGTGGAGCGGGGTGCGGGGCCACGG 101	
Db 34410	CGCCTGCATTCACAGTGGAGCGGGGTGCGGGGCCACGG 34450	
RESULT 2		
LOCUS	ARI72664	71989 bp DNA linear PAT 17-DEC-2001
DEFINITION	Sequence 2 from patent US 6303342.	
ACCESSION	ARI72664	
VERSION	ARI72664.1	GI:17912155
KEYWORDS	.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified.	
ADTHORS	1 (bases 1 to 71989)	
TITLE	Julien,B., Katz,L., Khosla,C. and Tang,L.	
JOURNAL	Recombinant methods and materials for producing epothilones C and D	
FEATURES	Patent: US 6303342-A 2 16-Oct-2001;	
SOURCE	Location/Qualifiers	
	1..71989	
	/organism="unknown"	
BASE COUNT	10108 a 23531 c 26617 g 11731 t 2 others	

ORIGIN	
Query Match	100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity	100.0%; Pred. No. 2,7e-15;
Matches 101; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GCCTTCGGGATGATATACCGCGGTGGCGGTGCAGTTCATATGCGACGACGAGACGACCGACG 60
Db	34350 GCCTTCGGGATGATATACCGCGGTGGCGGTGCAGTTCATATGCGACGACGAGACGACGACG 34409
OY	61 CGCCTGCAATTCACGTTGCGAGCGCGGGGCGCGGCGACG 101
Db	34410 CGCCTGCAATTCACGTTGCGAGCGCGGGGCGCGGCGACG 34450
RESULT 3	
AF210843	
LOCUS	AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION	Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
ACCESSION	AF210843
VERSION	AF210843.1
KEYWORDS	GI:6724237
SOURCE	
ORGANISM	polyangium cellulosum. polyangium cellulosum Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium. 1 (bases 1 to 68750) Molnar,I., Schupp,T., Ono,M., Zikkle,R., Milanow,M., Norak-Thompson,B., Engel,N., Toupet,C., Struam,M., Cyr,D.D., Gottlieb,D., Mayo,J.M., Hu,A., Golf,S., Schmid,J. and Ligon,J.M. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum so ce90 Chem. Biol. 7 (2), 97-109 (2000)
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
FEATURES	
source	Location/Qualifiers 1..68750 /organism="Polyangium cellulosum" /strain="So ce90" /db_xref="taxon:56" c1..1826 /note="ORF1" /codon_start=3 /transl_table=1 /product="unknown" /protein_id="AAF26904.1" /db_xref="GI:6724238" /transl_table=1 /note="ORF2" /codon_start=1 /transl_table=1 /product="unknown" /protein_id="AAF26905.1"

/db_xref="GI:6724239"
/translation="MGALISVAPAGCALGAEDEEGOPGODAGAGALAPAREVMAEVA
ACMPGAVMLVARGDVAVGATVETLGGSPAPMRDITIFRIASMTKAVATAMVME
EGKLDSDEVDRLPELANRKVYRIDGPIDETVEPAREITVDELMTFMGESISDA
SSPIORADELGLVNAOPYPMTPHGPDEMRRLGCTPLMHOBQACQMYNMGESIVC
VGRADQGFDAVYRERILAPLGRDITDIFVPADKLARFACCGFTTBQGTGKTRMD
GABSAVASPPAPSGAAGLVSTVDDYLLFRMLMNGVHEGRLLASAVREMTADHL
TPAKAASSEFPFQETTHGWCMAVVTAPDANSEVPGRYGMDGFGTSMINDPGREL
IGIVMTQSAGFLFSGALERFRMSVYVATESA"
3415..5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGLTEROVLISVLTALITVYARASGELARLRDPEVLELFG
GVYLGPSVYGALAPGRHAFQEPANGVLSGISMGALLLMLAGIEVDGLIKREA
RPGALSALGAIAPPLAAGAAFSALVLDRLPSGLFGLIVSVYAVIAKVLIERESM
RSTYQVTLAAGVSEVAAVLAAMTSSSYGASPALAVARSALLASGLFLFVILGRS
LTHLAMRVADATRVSKGVSLVLTFLAALTQRLGHLPLAGFALGVLAKVAPRT
NRPDLGVOTLVAGLEAPVEFVLAGRVVSOQLRTPAANGVALLATATPAKKVVA
IGARLGGLRSEALVAVGLNMGCTDLVATVYFGLGSLNAYMYVAVLTATYTA
SPALLIMLEKRAPPQTESARLREBAARATIPCYERILVPIVAHALGFRDILES
IVASKRKLGETVDITELSVEQAPGPSRAGEASRLARLGRVIGLRQRELIGES
IOAILASRDHDLVIGARSPARAGMSFGLQDAIVORAESNVLYVDDPDAERAS
ARRILVPIGLEYSFAADLAAHVALAMDAEVLSSAQTDPCA VYMRDREPSRVAV
ARSVDVEAFGRGLRGVRSRYHVGAFHSPDETRILARAPDLVLGVGDHGPGLRL
YLGSTVEYSVRVPAVLVAHNGTRQGRV"
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGTRCAIAESYIPGSGNPPQMTSHTACLLASORDA
QVAITYFSDRDPAGPYRVTVPARRTRHVRFDLTPEDPIPRDTVASYIESDAPIV
QRLRLDSROAENALSTIAYDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVLTGRFRGRHRIGHTIGLISMAALAGGSPSEKTYOCTRIAP
GADARTADVDDAATTRLAVDVYHLSPPERLEAGSERVYWMORPSPEPMRWVGLD
YNADSRKGLAETVYANFELITAEKQSSPQSPSSAAVIGTFSVG"
7610..11875
/gene="epoa"
7610..11875
/gene="epoa"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRIERAEDPIAVGASCRLPQGVLDISGFWTLLGSSRPTV
GVPAREMDAAAMPDPDPAGKTPVTRASFSLSDVACFDASFFGISPRALNMDPAHR
LLEVCWEALENNAIAPSAVLGETGVFFIGIPSEVLAALPQATSAEIDAGLGTMT
PSVAGRI SYALGRLPQVAVATVAVSSSLVAHLACQSLRGSCSTALAGVSLMSP
STLWLSKTRALARDGRCKAFSAEADGFGRGCAVYVLRKLSGARADDDRIILAVRG
SAIINHGAASSGLTVPRGSSQSEYLRKALADACCAASSVGYEAKHGCTTLLQIEQA
LNAVIGLGRDVAFTPLIGSVKTNLGHPEYASGITGLKVLSLQHQIPAHILAAQNL
PRI SMGDLRLVYTRARTPMDNTPRRAGVSSFGMSGTNAHVYLEAPATCTPPAPE
RPAELVLSARTASALDAQARLRLDLETYPQCGDVAFLATTRSAHEHRLAVAT
SREGRAALDAAGOTSPGAVRSTADSSRGLAFLFTGOGAOTLGMGGVLDWMS
REAPDLCVLFTNOLDRPLREYMAEPASVDAALLDOTAFQOPALFTFPYALAAALRS
KQVEBELVAGHSIGELVAVACVAVSFLEDAVFLVARGRLMOLALPGGAMVSTEAE
DVAAAVAPHAASVSLAAVNAPOQVYIAGAGQVHAIAAAMAAARGARTKLAHSHARS
PLMAADLFAFGRAVSYSRRPSIVLSNLSGKACTIDEVSSPGYWRHARREYVREFADG

Query Match

misc_feature
/note="Region: beta-ketoacyl synthase/malonyl-CoA decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P: contains condensation/heterocyclization, cysteine adenylation, putative NAD(P)H oxidase, and peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQLNELEHGKIKLADEERLQIQAPKNALNPILARISEH
KSTILMLRQRLPAESIVAPAREHAPFELTIOESYMGRTGATVPSGIAHYREYD
CNDIDVRLSRARFKVYARHMLRAHTLPDMQIOLEPKVADIEITDLRGLDRSREA
RLVSLDAMASHRIYDTERPRLHVAVARDEQOTRVLSDILINDIGLSITFFKWL
SFYEDPETSPLVLELSYRDVLAESRKSSEAHOSMOWKRIARIELPRLPKAD
PSTKTRPRHTEOMLPDSKGLKRVGECRLTPGYTLAAFSYETIGKWSAPRFL
NITLENRLPVRPVRNDITGDFTSWULDITDTRDSFEQRARIDEQULEAMDHQDVS
GIEVOREARVVLGIONGALPEVLTLSALNOQVGYTSIORLGTPLYTSTQTPOLLIDH
QYEHGDDVLVAMDIVDGVPDPDLDDMLEAVYARLTERPEMGQVRCSLPPAQLE
ARASANATNALISEHTLHGLFAPARVQOLMOAVYASRKTLYEELSRSRGLAR
EGARPNITLVAAYMEKMGQVAVAVLIESGAAYVPIODILPAEBILHVLHDGEVXIV
LTQPIUDGLKSNPPGIGIQLVSEAGVEGDGDPMPPIOTPPSDLAIVYITSGSLPK
GVNIDRGAVNITLIDINERFETGPDQVLTALSLSFDLSVYDFGLAAGTIVVDA
SKLRDPAHAEELIEREKVYVMSPALMRLMVEHGRDLSLRSLSLGDWIPV
GLPGELOAIRPCEVSYISLGATASIMSGYVRVNDLSMASIPYGRPLRNQTFVLD
EALPEPVPVPGOLYIGVGLAGVWDEEKRKSLVHPERGERLYRTGDLGRYLD
GNIEPMGRDNQIKLRGYVELGETIEETLKSHPNADVIVPGDANKILLAAVVP
EGTRRAAQODASLTERIDARAHAADGLSDGERVQKRLARHGLRDLDPVVD
TGQDPREAGLDVYARRRSVTRLEAPITPEVEGRFLSCISVSVEGATLPKRYPSAG
SYTPQVOTVYVRSGLIEGVDEGFYVHPEHRLKLSDIGIEGRGAHVRENDVEDEAA
FNLTFVGRIDAIESLYGSSREFCLLEAGYQMLLEQAPSNICVCEVGOQNFQOVR
PVLDLRHSDVYVHMGSRVDRPOVOCGLGQDSSRRATTCGAPGRGROMLNLRD
FLRTKLPEYMPVPEVLEALPLTSGNKYDRALRERDTSPPRSGHAPRPALEET
LVAVREVIGLEVYGLQDSFVLDGATSHIYMRSLQKRLDRELAITTELFQYMLGS
LASGLRDSRDLDQRPNMQDRVEVRKGRRS"
16251..21749
/gene="epob"
16251..21749
/gene="epob"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

98.4%; Score 99.4; DB 1; Length 68750;

Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 40021
|||||

OY 61 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 40062
|||||

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 40021
|||||

OY 61 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 40062
|||||

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 60
|||||
Db 40022 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 40062
|||||

Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 40021
OY 61 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 40062
|||||

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 40021
|||||

OY 61 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 40062
|||||

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 60
|||||
Db 39962 GCCTTCGGGATGATACCGCGGTGCGGTCATGCGCAGGAGGACGACG 40021
|||||

OY 61 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCTGCAATTCCACGTTGCGAGCGGGTGCCGGCCACGG 40062
|||||

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
JOURNAL Genes for the biosynthesis of epothiliones
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6,7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GCCTTGGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACGAGAGACGACGAGC 60
|||||
Db 39962 GCCTTGGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACGAGAGACGACGAGC 60
|||||
OY 61 CGCCTGCAATTCCACGTTGGGAGCCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCCACGTTGGGAGCCGGGTGCCGGCCACGG 101
|||||
RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
JOURNAL Genes for the biosynthesis of epothiliones
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6,7e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GCCTTGGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACGAGAGACGACGAGC 60
|||||
Db 39962 GCCTTGGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACGAGAGACGACGAGC 60
|||||
OY 61 CGCCTGCAATTCCACGTTGGGAGCCGGGTGCCGGCCACGG 101
|||||
Db 40022 CGCCTGCAATTCCACGTTGGGAGCCGGGTGCCGGCCACGG 101
|||||
RESULT 10
LOCUS AX024384 16124 bp DNA linear BCT 15-SEP-2000
AX024384

DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES
source Location/Qualifiers
1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 64.6%; Score 65.2; DB 1; Length 16124;
Best Local Similarity 80.9%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
OY 2 CCTTCCGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACCGGAGCGACCGAGCC 61
|||||
Db 2638 CCTTCCGCGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACCGGAGCGAGC 61
|||||
OY 62 GCCTGCAATTCCACGTTGGGAGCCGGGTGCCGGCCACGG 95
|||||
Db 2698 GCCTGCGGTTCCAGTACGAGTCGGGAGCCGGG 2731
|||||
RESULT 11
LOCUS AX024277 16124 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
REFERENCE 1 (bases 1 to 16124)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES
source Location/Qualifiers
1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 64.6%; Score 65.2; DB 1; Length 16124;
Best Local Similarity 80.9%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
OY 2 CCTTCCGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACCGGAGCGACCGAGCC 61
|||||
Db 2638 CCTTCCGCGCGATGATACGGCGGTGGCGGTGCAGTGCATGGCGACCGGAGCGAGC 61
|||||
OY 62 GCCTGCAATTCCACGTTGGGAGCCGGGTGCCGGCCACGG 95
|||||
Db 2698 GCCTGCGGTTCCAGTACGAGTCGGGAGCCGGG 2731
|||||
RESULT 12
LOCUS AY112896 609 bp mRNA linear MM 09-JUN-2002
DEFINITION Equus caballus Indian hedgehog (IHH) mRNA, partial cds.
ACCESSION AY112896
VERSION AY112896.1 GI:21360570
AY112896

SOURCE		horse.
ORGANISM		Equus caballus
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus. 1 (bases 1 to 609)
AUTHORS		Semevolos,S.A. and Nixon,A.J.
TITLE		Partial nucleotide sequence for equine Indian hedgehog mRNA Unpublished
JOURNAL		2 (bases 1 to 609) Nixon,A.J. and Semevolos,S.A.
REFERENCE		Direct Submission
AUTHORS		Submitted (16-MAY-2002) Clinical Sciences, Cornell University, VMC C3-176, Ithaca, NY 14853, USA
TITLE		Location/Qualifiers
JOURNAL		1..609
FEATURES		/organism="Equus caballus" /db_xref="taxon:9796"
SOURCE		<1..>609
gene		/gene="IHH"
CDS		<1..>609 /gene="IHH"
		/codon_start=2 /product="Indian hedgehog"
		/protein_id="AA04071.1"
		/db_xref="GI:21360571"
		/translation="RNRVGLLARLAVAGEIDWYVESKRAHYHCSVKESHSAAKTGGCC EPGAQVRLSESGARVALSAVPGRVLAMGDEGPPTSDVLIEDRPDLRALAOVIEA TDPPRLALTPAILFTANNHSEPAHFRTATFSOVPQGVYLVAAGVGLQPARVAIA VSHTVALGAPAFIRHGCTILVEVDVASCFAFVADHIQLAE"
BASE COUNT		102 a 195 c 196 g 116 t
ORIGIN		
Query Match		34.1%; Score 34.4; DB 4; Length 609;
Best Local Similarity		63.1%; Pred. No. 90;
Matches		53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY	6	GCGGANGATACGCCCGGTGCCTCATGCGACGAGAGCACGACGCCGCCG 65
Dn	271	CGAGCGGATGATGAGGTCGAGGCGCTTCCAGTGCAATTGAGACCAGCCCCGCGCGCT 330
OY	66	GCAATTCACAGCTGGGAGACCGCGGT 89
Dn	331	GGCACCTACGCGCTGCCACACTGCT 354
RESULT 13		
AE012031		
LOCUS		Xanthomonas axonopodis pv. citri str. 306, section 409 of 469 of the complete genome. AE012031 AE008923
DEFINITION		AE012031.1 GI:21110216
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		Xanthomonas axonopodis pv. citri str. 306. Xanthomonas axonopodis pv. citri str. 306 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.
ORGANISM		
REFERENCE		1 (bases 1 to 10507)
AUTHORS		da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furian,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergo,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L., Cunha-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,N.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kish,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locall,E.C., Machado,M.A., Medeiros,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,Y.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,A.D., Silva,C., de Souza,R.F., Spinola,I.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities	Nature 417 (6887), 459-463 (2002)	22022145	2 (bases 1 to 10507)	da Silva, A.C.R., Monteiro, J.A., Retnach, F.C., Farah, C.S., Furian, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannan, F., Cardoso, J., Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorri, H., Farla, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsiyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, V.F., Locati, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.
Direct Submission	Submitted (28-NOV-2001)	Departamento de Biogenética, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900, Brazil			
FEATURES	Source	Location/Qualifiers			
gene	source	1..10507			
gene	source	/organism="Xanthomonas axonopodis pv. citri str. 306"			
gene	source	/strain="306"			
gene	source	/db_xref="taxon:190486"			
gene	source	/note="pathovar: citri"			
gene	source	complement(152..775)			
gene	source	/gene="gst"			
gene	source	/note="XAC3819"			
gene	source	complement(152..775)			
gene	source	/gene="gst"			
gene	source	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/genemark"			
gene	source	/codon_start=1			
gene	source	/transl_table=1			
gene	source	/product="glutathione S-transferase"			
gene	source	/protein_id="AA38661.1"			
gene	source	/db_xref="GI:21110217"			
gene	source	/translation="MPAITTGRRSSNSVKKAIKACAEAGVAVYSIEVGAGGNDP ATRLNNGVPTLQDGLVLMESNATVRLIAQVAPALPQAPAEALGDWRMDPTT STFGFRDLFWGVRTPEARHARIALAALQSGELARDAALAOQPLISGGRFAM GDLPISFIYAEEMPIERLEPHLQAWYARLRVAPYORGVMYAL"			
gene	source	complement(811..1617)			
gene	source	/gene="XAC3820"			
gene	source	complement(811..1617)			
gene	source	/gene="XAC3820"			
gene	source	complement(811..1617)			
gene	source	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/genemark"			
gene	source	/codon_start=1			
gene	source	/transl_table=1			
gene	source	/product="conserved hypothetical protein"			
gene	source	/protein_id="AA38662.1"			
gene	source	/db_xref="GI:21110218"			
gene	source	/translation="MSMSDNDARIALIDADNAPACKIDVLAEVARYGVANRRVY GNKSPILKGEALHDYAIRPQIOFAYSSGKNSDAMVYIDAMDLYARLDGFATG SSDDPFLWRLITDGMKVGEGEETPAFVAVGSKFTYVLEIGDAAAASPTAG RKQATAIRDTRLYOMLRNALVSCEDDGALISAVKQVANGASPPRYNGYRKSLD LVRIATGEIRQDQALWVRTPKSGGKPPQAPAAARPAAPAKPPAAG"			
gene	source	complement(1614..2204)			
gene	source	/gene="XAC3821"			
gene	source	complement(1614..2204)			
gene	source	/gene="XAC3821"			
gene	source	/note="putative; ORF located using Glimmer/genemark"			
gene	source	/codon_start=1			
gene	source	/transl_table=1			

/product="conserved hypothetical protein"
 /protein_id="AAC38663.1"
 /db_xref="GI:21110219"
 /translation="MSAVRHHNVASDASGHPNSSLPPPLPLSSVPPAAPSRAAYVVK
 KPLSSVALGLGVATCYVLRGTVLGLALIMLSIGSLPADARSALQADAVYENHIG
 IIRHQVDLYLSTSLAPSDDSEDFVDEGDDGTGTVATWSDAREILISTGVLLRGGQ
 TEYLPATVDVEEEMDTDTDTDFDKTESAKSLQG"
 /length=3561
 /gene="ndh"
 /note="XAC3822"
 /note="ndh"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="NADH dehydrogenase"
 /protein_id="AAC38664.1"
 /db_xref="GI:21110220"
 /translation="MSTSLPSQAPALILVYGGGPGVGMATRALDPGICITLVLDQNHGHTL
 LPQLIYVATAGISAPDIAAPRIHLRDRVREYVLGDVDTIAPSRREYVLAQGGHTL
 GYDMLLATGTGTAHYFENDMDAAHAGLTKTLLDVALYLRKLLLEEREAASDPAARA
 AMLSAAVVGGGPPEVETAGTLAGTARIHTLKNERRHIDPQARVRLVEAGPVLSPFEE
 DLTAKARLQELERLEVENGTPTQDIALGAYGLGDFEYVAFVVAAGVASSPLARLL
 GVPIDRAGRVLYEVRSDSEVPHPEFVGGDASVQODGPRVPEVPAAKOMKHAKIAI
 BARQGVTAARVPRDQDGNATITGRMAALVYHVKIKLSGVAAVMFWMLAHYFLIGERR
 NREYVLMMAMATVSTQRRARITLFGGATDDPPSSKDG"
 /length=3614
 /gene="XAC3823"
 /note="XAC3823"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC38665.1"
 /db_xref="GI:21110221"
 /translation="MATVLELLDLLETFPALSGATVAVNRRLDFGVLYSCAASV
 GGIVATVIGATVPAALVQVYLIVLVCAGVDFGCHAEVLRNPVIFPAAGLAF
 AVYGSADLIDHPLPSALITLMLSGIGGIRADLVARTPVYLQAEILYAAALAGG
 LVAVGHVLDVQAPASLTAGAGVCFGLRFMAIRYGMHLPVARSPE"
 /length=4307
 /gene="rpoH"
 /note="XAC3824"
 /note="XAC3824"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase sigma-32 factor"
 /protein_id="AAC38666.1"
 /db_xref="GI:21110222"
 /translation="MNTQSTFALVANNLPIPSALGSIAVYTGAVVQHVQSVDEEDDL
 ARRRDELDIDAAEYLHSLRRVYVYAGSYNGTCLPGLDIOEGNITGLMAKYRFPD
 DMGAKLVSAFVAHIVRAIRAEHFIILKNRIVKVAITYAQRLEFNLSKSTRGLINAAE
 TVAAKGLVASERMVMEBSRLSGRIDASDEDDHGPSPVSYLVANEEDPSSQA
 YERHDSDENQQLILREEMAGLDRSDIVKRRWLDSKSVTLQELADEYGSARIRIQ
 IEANALKNQLLFFVA"
 /length=5381
 /gene="ung"
 /note="XAC3825"
 /note="XAC3825"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="uracil-DNA glycosylase"
 /protein_id="AAC38667.1"
 /db_xref="GI:21110223"
 /translation="MTEGSRIDLEPSMKARVGDMLLRPQMRSLSATLRQKKAAGAVV

gene
 complement(6103. .6897)
 /gene="XAC3826"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/genemark"
 /codon_start=1
 /transl_table=1
 /protein_id="AA38668.1"
 /db_xref="GI:21110224"
 /translation="MYYDGSKIVRKLIADVLKRLPNVQYIGCSNIEARQALEAGAV
 DLVTLISLDPDGDLTLARSVRETAGAYPVVIVSDAQOHLVERFTEYVDYFQK
 ALGHEALATEIRNGYQPEPVVGATVLIEDSRVAEATKRLKSNLVVHLVLAEDPA
 FALTALSLGTERRIDVILVDLKEGLNGRDVEVERLIDFAGKRLPVLVMTGQGT
 NPNRQSDLLRGANDLVOKPIEERLIVTKVLFGURLARLEDAQVTL"
 complement(6954. .7904)
 /gene="ftsx"
 /note="XAC3827"
 complement(6954. .7904)
 /gene="ftsx"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/genemark"
 /codon_start=1
 /transl_table=1

Query Match	Similarity	31.5%	Score 31.8	DB 1	Length 10507
Best Local	Similarity	57.6%	Pred. No. 2,8e+02		
Matches	57	Conservative	0	Mismatches	42
				Indels	0
				Gaps	0
OY	1	GCCTTCGCGATGATACGCCGGTGGCGGTGCAGCTCATATGCGAGACCGAGAGCACCAGC	60		
Db	3061	GCCGCGCTGTGGCGCGCTCCCGCTGGCGGTACAGCTGGGGCGTGCCTCGCATTCGCCCGCGC	3120		
OY	61	GCGCTGCAATCCACCTGTGGCGAGCGCGGGTGGCGGGCGAC	99		
Db	3121	GCGCTGCTGTGAGAACCGACCTCAGCTGCGCGGGCCAC	3159		
RESULT 14	AE007185/c				
LOCUS	AE007185	17205 bp	DNA	linear	BCT 27-APR-2001
DEFINITION	Mycobacterium tuberculosis CDC1551, section 271 of 280 of the				
ACCESSION	AE007185				
VERSION	AE000516				
KEYWORDS	AE007185.1 GI:13883793				
SOURCE					
ORGANISM	Mycobacterium tuberculosis CDC1551.				
	Mycobacterium tuberculosis CDC1551.				
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
	Actinomycetales; Corynebacterineae; Mycobacteriaceae;				
REFERENCE	Mycobacterium tuberculosis complex.				
	1 (bases 1 to 17205)				
	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,				
	Peterson, J., Deboy, R., Dodson, R., Gilm, M., Haft, D., Hickey, E.,				
	Kolony, J.F., Nelson, W.C., Umayan, L.A., Ermolaeva, M.,				
	Salzberg, S.L., Delcher, A., Uitterback, T., Weidman, J., Khouri, H.,				
	Gill, J., Mikula, A. and Bishai, W.				
REFERENCE	Whole genome comparison of Mycobacterium tuberculosis clinical and				
	laboratory strains				
	unpublished				
	2 (bases 1 to 17205)				
	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,				
	Peterson, J., Deboy, R., Dodson, R., Gilm, M., Haft, D., Hickey, E.,				
	Kolony, J.F., Nelson, W.C., Umayan, L.A., Ermolaeva, M.,				
	Salzberg, S.L., Delcher, A., Uitterback, T., Weidman, J., Khouri, H.,				
	Gill, J., Mikula, A. and Bishai, W.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712				
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					

Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
source
1. 17205
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db.xref="taxon:83331"
/note="clinical strain"
complement(130. 1062)
/gene="MT3910"
complement(130. 1062)
/gene="MT3910"
/note="similar to PID:804884; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="esterase, putative"
/protein_id="AAK4827.1"
/db.xref="GI:13883794"
/translation="MPTALHDGGHMKRSALLRALMIALSFGIGVAVAAEPTAKA
APYENLVPSPMGRDIPVAFIAGGPHAVYLLDFNAGPVSVMVNTGNMNTLAGG
ISVAPAGAVSYMTNEDGSKOMDFLSAELPDMIAANRGLAPGHAIVGAAGCY
GAMALAFHDFRFGAGSMGSLTPSTTTNGAIAAGMOQGGVDITGMGAPQLGFW
KMHDPVHASLLAQNNTRVWWSPTNPASDPAMIGQAAEAMGNSRMPFYQYRSVG
HNGHDFPAGSDMGSMAPQLGAMSGDIVAIR"
complement(1209. 2276)
/gene="MT3911"
complement(1209. 2276)
/gene="MT3911"
/note="similar to SP:P31952 GB:X63398 PID:44564
PID:1491793 PID:225974; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="esterase, putative, antigen 85-A"
/protein_id="AAK4827.1"
/db.xref="GI:13883795"
/translation="MRPMHGCVMRREARERAMOLVDRGAVTGMRSRLVYGAVAL
VSGIVGVTATAGATAGSRPLPEYELQVPSMGRDITKVOFOSGASNPALYDGL
RADDESGMDINTPAFEMVYDQSGLSVMPVGGGSSFTSDYOPACGACGCTYKWEF
LTSLEPGLDANRHVKFTGSAVVGLSMAASSALTALYHPQOFYAGANGSLDPSA
MGPTLIGLAWDGGYKASDMGKEDPAMORNDPLINVKLIANNRVWYCGNGRP
SDLGNNLPARKFLEGFVRTSNIKFQDAYVNGGCHGVDFPDSGTHSMIYGAQLNAM
KPDLRALGATPNTGAPDGA"
complement(2520. 4445)
/gene="MT3912"
complement(2520. 4445)
/gene="MT3912"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4827.1"
/db.xref="GI:13883796"
/translation="MVGGRVPVPEPYRVVSVLSMTSVTAVALFGWGSNORWIAAD
GLIVRTVRLNLAGNGPVFNOGEREANTSTAMTYLLYVGWVGWGPATSLGESLIVA
VLSILGVLMLIGRLYAPLSGRRAIMLPAGALYIAVPADPAPFSGESLIVA
YLGILMMMYCWSOPLRARPSQMLGALAFVAGCSYLVPERALLIGLALIMLILAA
RTMRRLVLYLAGGFVAVOIFRMGTIGLLVSTALAKDAAGKWSQGMVYSNER
PYALWPLVLSVPLGLLMTARRPSFLRVLAADRVARAAGSPRAVAFYVGSV
LOALYVIRGGDFWGHVLLAPLECLAPGVILPDPKDSRETRGMLVGLISGL
WLGJAGSLMAANSFGMDATRYVTSIYDEREYQAOTGAHAPLTAADLYDPRA
AVLTANNPEGALLPLSGYNOMDIPMTIPSSGTAPGKPAKPOHAAFEPTMGIL
GMVGLDVRVYIDQIGLVNPLAHTERKAKRIGDKRLFPDMVYADSPWVKYFGITG
YIDQVYTOABALQCPATRAVLNSVAPITLHRELSNVLSHSEYKTRIDRVPRYL
VRCGLVDPGPGPPRE"
complement(4492. 5400)
/gene="MT3913"
complement(4492. 5400)
/gene="MT3913"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

/protein_id="AAK4827.1"
/db.xref="GI:13883797"
/translation="MSPDVYTOPANLVAGVYAIRPROMKKNVLIAPLALGGV
RTDYEVLSKYSMAFVFSLAASAVLYLVNDYRDYEADEHREPTKRFRIAGVPEMLA
YVAAVLVGTVSLAGAMMLTPNLALVMVYLAQMLAYCEGLKHOAVNLCVSSAYTLR
AIAGVATKIPLSKMFLLIMAFSLFVPAKRAELHARTGAIAIKSSLESTYSL
RFVWTLSTAVILCYGLMAFERDYSGSMAVSMIPTTIALIRYAVDVGGLAGEPD
IALDRVYLOLLALAMVATVGAAYFG"
complement(5407. 5904)
/gene="MT3914"
complement(5407. 5904)
/gene="MT3914"
/note="identified by match to PFM protein family HM
PF01569"
/codon_start=1
/transl_table=11
/product="Pap superfamily protein"
/protein_id="AAK4828.1"
/db.xref="GI:13883798"
/translation="MVAVQSALVDRPGLATPARGLSHGEHCIGMLILALGATLPR
RRRWLYAGAGAFYAHAIATLIKRVRRORDRPAIVAVNDTPSQLSFPANHSTTA
AALMGRTGLPLPVLPVPPAUSRILIGVHYPSDVAVGALGATVGAIVDSGGHQ
RAKRR"
complement(5933. 7846)
/gene="MT3915"
complement(5933. 7846)
/gene="MT3915"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4828.1"
/db.xref="GI:13883799"
/translation="MSLSLASLSRVLLPRPGEVDRLKYLEESTNARNAPIRT
SLQIASEVSFAFYFNAPASVYRRTTKSVLRQVYAGAVDVRKARAGARTE
VEGHDFTEGDPAAVETEYVLIQFEEGGWVPEITDTAVTLHSGWATSPAGTA
NIAVGIPTFNPACVNAKRELTPVDVQIYVYDQGERKVRDHPPEAPAAAT
GSRSLTHDQDLSGGSGYSRWYEAALKTDCOQITLFPDDIRLEPDSILVIMHRA
KAPLVGQMLINDEPSHILIMGEVYDRSIFPMWTAHAIEDHDFATYPLINDNSRK
LKHRLIVDYNWWTCTMIPROVAELQPLPLETKWDADYGLAAAHGYPVTLPGA
AIWMAWSDKADIDMOWAYFLRNRLVVAAMHMGPAQYIGVRSILKATLKLACL
EYSTVAIONDAIDFLAGPEHISILSALPRVIRKSYDPAVYLAASELPPPLIK
NKAKPVPNPLVICYRLARGIMHNLTPANPOHHRPEPNVPTODARWFLCTQVATV
TTADGCGVVYRORDRAKMFALLMOSLSLRQQLKRFEMRIRYDALPTLSQKWET
ALLPAAQDEBEHG
complement(7843. 9042)
/gene="MT3916"
complement(7843. 9042)
/gene="MT3916"
/note="similar to SP:P37747 GB:U03041 GB:U09876 PID:508242
PID:510253; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="UDP-galactopyranose mutase"
/protein_id="AAK4828.1"
/db.xref="GI:13883800"
/translation="MQPMTARFDLPVYSGFPGITLIERVATQDKNVLYLERRPHG
GNAVSEAPQGTLEVHYGALHFTSKRWYDVKQFTDTDKHRYFAHNGCAVQF
PMGIGLSQFEGKFTPEQARQLAEQAAIDDAQNLBEKASLGRVLEAFVNG
YTAQWQTDKELEIPANITRLPYRTYTDNRFESQYGLPDGTATLQMMADHRIE
VLANTQEDVDRGAPGSPAPRVYVYTLRDPYAGRGMTLDEVDVPLIGDQ
GTAMVNNIDIVPRTIHERHPRPREDYPTDKYIMREYSRAEDDDEYYUPLNFA
DRALATYRRAKASSETASSVYLRGRLGTYOYLDHMAIASALMNDVNLAPHLRQV
PLUDGA"
9306. 10160
/gene="MT3917"
9306. 10160
/gene="MT3917"
/note="similar to PID:845583 GB:AL123456; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="exported repetitive protein"

```

gene
CDS
    /protein_id="AAK48283.1"
    /db_xref="GI:13883801"
    /translation="MPNRRRLSTAMSAVALAAVAPCAFLVYESTETTEREHE
    FKQAVLTDPELMSALSGLSGFIQNPVPSLIPNDLSDGLGRPAKE
    GLNPGPRSTDVFGVTTTQVAVATRPDADITQPPAGRPNDLSDGLGRPAKE
    OPGGNISAILSPQAPPGTQMTPTATVTAAGOPPIISBAEMGDESLRCEPEYD
    RGVRAVYHHTAGSNDYSPLEAGIYKATVYHSHKIGCDIAYANLKYGVCEGS
    AGGLTRPVGEFTTGGFNRTMYAMIGNDDVAPPTQITRTYGRLLGRKGDVDPK
    SNVDLQASGSIYTFPGGAIARLPALFTHRDVGNTDCPNAAVAVWDETRDIAHFND
    PPEELIKALEGGAITYQWQALGMSNLSALCAPSPENDADGARYATFAKAMVWSPVT

Query Match      31.3%; Score 31.6; DB 1; Length 17205;
Best Local Similarity 58.5%; Pred. No. 3e+02;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

```

QY 1 GCCTTCGGGATGATACGCCGGTGGCGGTGCAGTCATGCGACGAGCAGCAGCCAGC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6505 GCCTGCTGGATATATGACGATGCGATCCAGACAGCAGCATCTTCGCGC 6446
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 CGCCTGCAATTCACGCTGGAGCCGGGTCGGC 94
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6445 GGCCCGGACGACATCTTTCGATCTCGAATCGG 6412
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 15
MTV026/c      23740 bp      DNA      linear      BCT 03-AUG-2001
LOCUS      Mycobacterium tuberculosis H37Rv complete genome; segment 157/162.
ACCESSION      AL022076 AL123456
VERSION      AL022076.1 GI:3256026
KEYWORDS
SOURCE
ORGANISM      Mycobacterium tuberculosis H37Rv.
                Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
                Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
                Mycobacterium; Mycobacterium tuberculosis complex.
                1 (bases 1 to 23740)
                Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
                Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
                Tekla,E., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
                Connor,N., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
                Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLaren,J.,
                Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
                Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
                Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
                Barrall,B.G.
                Deciphering the biology of Mycobacterium tuberculosis from the
                complete genome sequence
                Nature 393 (6685), 537-544 (1998)

```

```

REFERENCE
AUTHORS      Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
                Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
                Tekla,E., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
                Connor,N., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
                Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLaren,J.,
                Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
                Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
                Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
                Barrall,B.G.
TITLE      Deciphering the biology of Mycobacterium tuberculosis from the
                complete genome sequence
JOURNAL      Nature 393 (6685), 537-544 (1998)
MEDLINE      98295987
PUBMED      9634230
                2 (bases 1 to 23740)
REFERENCE
AUTHORS      Parkhill,J.
TITLE      Direct SubMISSION
JOURNAL      Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
                tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
                Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de GenetiQue

```

COMMENT

Molecular Bacteriome, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk

On Jun 26, 1998 this sequence version replaced gi:2961395.

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.

(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.

Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

```

source
misc-feature
RBS
gene
CDS
    Location/Qualifiers
        1..23740
        /organism="Mycobacterium tuberculosis H37Rv"
        /strain="H37Rv"
        /db_xref="taxon:83332"
        <1..>23740
        /note="fragment designated v026. Does not represent a physical clone"
        60..64
        /note="possible RBS for Rv3797"
        72..1853
        /gene="fadE35"
        72..1853
        /gene="fadE35"
        /note="fadE35"
        /note="Rv3797, (MTV026.02), len: 593. fadE35, Probable acyl Co-A dehydrogenase similar to E. coli A1dB protein, gpL20915|EC0A1DB_3 E. coli (A1dB) gene, comp (541 aa); FASTA scores: opt: 455 z-score: 561.5 E(1): 9.2e-24, 31.18 identity in 514 aa overlap. Also similar to MTG176.2 M. tuberculosis cosmid (542 aa), A1dB, 31.9% identity in 458 aa overlap"
        /codon_start=1
        /transl_table=11
        /product="fadE35"
        /protein_id="CA11861.1"
        /db_xref="GI:2950416"
        /db_xref="SPTREMBL:O53377"
        /translation="MPYDLEAVDKLPSTPEKQKQRYOTENRGAMGNIWYITPPTQ
        ITMATILRPDELATFAEPHUTRIIGELTGCPVTRMAETDRNPRERIDRWGHDISRY
        LPESFIQSKRAVIEARQAVRDAAAGVPSIALFAADYDLDLQAGALATGNN
        VNSLTAYAPDPVREFVILKNSGEMDEAQLTERAGSDGLALETATRGSDVWL
        LNSFWFASNCNGAEVFLVILKNSGEMDEAQLTERAGSDGLALETATRGSDVWL
        GTRSAVGEIEFVDAEFLLSGEPADSGVAGFLVTRRRDSRNCVRIRRLDKL
        BRAVESICVAGORARAGALIDKPIMRKRIAEWVYDEALAVDFGCAANRQPR
        CLPQRIAYPTKTKCRGITYASDAIEIHGNGIETWPARLRDQAVNTIEGPD
        NTLCDVRGIEQTRAHETLRLDAVSVDDDTTTLVSRRLIEDDAATATWTKLD
        ROLAEARLEPPLAOFEMVDYAGALATEQAAWRAVRGTRKALVARLVARVLADQGL
        RGIIDDCDEALORFDELVAAGATAEQT"
        1980..3314
        /gene="Rv3798"
        1980..3314
        /gene="Rv3798"
        /note="Rv3798, (MTV026.03), len: 444. Putative transposase identical to IS1557 MTCY373_32, sp|O10621|Y02X_MYCTU_HYPOTHEICAL_50.2_KD PROTEIN CY37 (444 aa), 99.8% identity in 444aa overlap and similar to many, e.g. gp|U57649|DBU57649_1 Dibenzo(furan-degrading bacterium (163 aa), FASTA scores: opt: 767 z-score: 9.6e.8 E(1): 0, 67.3% identity in 168 aa overlap, or to BPIS1001.1 B.parapertussis insertion sequenc (406 aa), 24.9% identity in 402 aa overlap."
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein Rv3798"

```

```
/protein_id="CAA1863.1"
/db_xref="GI:2950417"
/translation="MNRVRLRLGLVDRKTVIEDIEFEEDDADGAGVIAVRPRSA
VLRRCGRGKRAAWYDRGAGLRKWRSLDWGIVEVLEAEARVNCPTHPGTVAVPAA
RHHGHGTAAPDVAAMLAAGVSKTAVCELMRIARVTGAIAVARADTEKRIIDFANL
RRIIDETISYKRHHRYLTVVVDHDSGRLVMAAPGHDKATGLDFEDALGAEEAAITVY
SADADMLADVAVFERCPDAIOGADPFVMAVMAPEALDVRERRANDARAIARPEKKG
RGRRKNAAARPRGERRARLRKAGVAAAKMKNPEDITERQSAKLMIATDRLRYAILL
KESLRHVSFKGEGKQALDWISMAQRCRIPVVELARIKRHRVAIDALDGLLSQ
GLESTMTKIRLRLIRIAFGFRSPQALIALMLVLTAHGRPTLGRHNPISQ"
/length=331
/gene="Rv3798"
/notes="IS1557-3rd copy"
/complement(3382..4935)
/gene="accD4"
/complement(<3382..4935)
/gene="accD4"
/notes="Rv3799c, (MTV026.04c), accD4, len: 517. Probable
propionyl-CoA carboxylase: Beta chain, similar to many,
e.g. SPIP53003|PCCB_SACER PROPIONYL-COA CARBOXYLASE BETA
CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E():
0; 50.7% identity in 519 aa overlap, Also similar to
U00012.17 M. leprae cosmid B1308 (549 aa), 49.8% identity
in 514 aa overlap and to MTCY1.20 M. tuberculosis cosmid
(548 aa), 48.9% identity in 515 aa overlap"
/codon_start=1
/transl_table=1
/product="accD4"
/protein_id="CAA1863.1"
/db_xref="GI:2950418"
/db_xref="SPTRMBL:053578"
/translation="MTYTPVYLHTTAELRLRLAKEPGEKAARDKGIPS
ARARIELVDPGSMETGALCRTPGDNALIGDVSIVGHGILNRPVGEISDDTVG
GTVGEMFGRVVARLMEWCAMVGCPIVGINDSGARIDAVTSLAMVLEGRHDLG
LVPOISITILKCGAGAVYSPIDLVVAVRPGAFVTPGVNGLVGEVSLDELG
ADHQASGINHGVYSEAAVQYVRDLFLPSCFPCPVNGLPEIIGSHLELD
SIVDSDNMAYDMHEVLRITFEDGDFDVAAQAQALITTGQARDGTGVVAAQMPH
MSGADINEASDKARTRFSDADPIPVYFVGPVQVQENGINIKKGRLLYAV
VEADVPKVTITTRKSTGATVAGSKULADLNFAMTARIAVGAGGAAQLKKRP
DPNAPEQAQIRKSEVENYNLMATPWIAERGFIDAVIDPHETRLRKSMHLRDKQ
LMRVGRKRG"
/complement(4932..10133)
/gene="pks13"
/complement(4932..10133)
/gene="pks13"
/notes="Rv3800c, (MTV026.05c), len: 1733. Probable
polyketide synthase, similar to many, e.g.
gpiM63676|SERERYAA.1 S. erythraea first ORF of eryA gene,
involved in complex polyketide formation in erythromycin
biosynthesis. FASTA scores: opt: 1329 z-score: 1411.0 E():
0, 33.6% identity in 1232 aa overlap or to AF040570.9
Amycolatopsis mediterranei rifamycin biosynthetic gene
cluster (3413 aa), 34.3% identity in 1162 aa overlap.
Also highly similar to mycobacterial putative polyketide
synthase: MTCY38.19, U00023.3 M. leprae,
MTCY22610.2, MTCY19h9_2, U00023.6 M. leprae, contains
P500606 Beta-ketoacyl synthases active site."
/codon_start=1
/transl_table=1
/product="pks13"
/protein_id="CAA1864.1"
/db_xref="GI:2950419"
/db_xref="SPTRMBL:053579"
/translation="MADVAESQENAPAEERAEELVPEKROWLRNMVGAQKAPDSIDE
SVPAVELGSSRDVAVMAADIEDLTGTLVSAFAFHTIESLATRIIEGPEFTDLG
DDAEDMSRTGPAERVDIAIVGLSTRPEGMNTPQTMALLLEGKDTLDPGMRSEF
LEEPRLAARVAGARTBGYLRKDIKGFDSSEFAVAKTEADNITDPOQRMALELTPGALSH
ARIPASLSRQAVGVYIGSSNDYSPLAVSDPYAHRYAITGSSITANRVSFTYDF
HGRSVITIDTACSSSLVAIHQGVQALRNGEADVAVAGVNALITPMVTLGDEIGAVLA
PDGRIKFSFADADGTVRSIEGGMLVLKRVDDARDGDAITLAVIAGSAVNHDGRSGLI
APNDAQADVILRAYKDAGIDPRTVDYIEAGTGTIIGDPIEALGRVVGGRPADR
PALGAVKTVNGHLESAAASMAKVVIALQHDKLPISINPAGSPYITDPAMRLKMI
```

```
TTPTDMPRYGVALAGVSSPFGGANAHHVVREVLPRDVEYKEPEPEPERKAAEPAP
APTLGAHLRDEERGNITDSAVAEEPEDEPLPGTEERALRIKEALIEIAOQETATL
VPLVASAFUTSRKAAAALADLMQOSBEGQSSLESIGRSISRNHRSRAVYTAH
DEAIKGLRAVAGKQAPNVFSDPVTGTEGAGAGAGAGAGAGAGAGAGAGAGAGAG
IEKVDALVOBELGYSLLELIDDAQDQGIETTVTIPAIQALGELLRHGAKPAAYI
GOSIGEASAFVAGLSLRDARTRAICSRSHLMGSEAMLFGEYIRLMAIYEASDEIR
EVFSDPELDLEVYAAAPROYIGGPPPOVDAILARAEKGFARKFATKGSHTSOD
PLIGELTAELOGITPSTPTGISTVHEGRYIKRGGRPHDVEWKKGLRHSVFTHG
IRNAVDSGHTTFELARNPALMVALITTDAGIHDQILPTLARKDEYSWSTVA
QLYVGHDLDRILFSRASGPODYANIPTPERFKREHMLPAHFGSDSTYMPGTHVL
PDGRHWEYARPDNDVLAALVRAAAHVLDPDQITAEORAVAGDARIVTTTRHP
GGASVQVHARIDESFTLVYDALVSRASSEVLPTAVGAATAIYADGAPVAPETPAD
ADAFETLSDSLTRBYMPSGMTRWSPDSGETTAERGLVGSAMGEPPDLMPREVLI
GLDSIMAVRIKRVYEDPDLPIPTLTAVRANILYVKEILIEYAYEHREDEVQVJHEOK
TQTAETARAQAEELHGRVGTKEPYDEAGVAFSPONGEOPNRTGALNDVYPRRA
AERYTFAWALVTGKSPGGLFNEPLRLDDEAAKIAORLHRSRAGPTTAEVLTSSNI
EALADKVRITYLEAGQIDGFVFTLRAREAGKVFVFEHPSRGGSTVYVYEPDLGRPD
TPMGFEFVEGSIIEERAQOYVPKLIEMQGDGPLYVMSLGGVLAACALIGRLGD
VRFVGLIDAVRAGEIIPOTKEIRKRMDRVAAAFKTEFNVTIPAIPEOLEELDECO
VRFVLDVAVSOSGVQIPAGIIEHORTSYLDNRAIDTAQIOPYDGHVTLVMDRYHDAI
MFEPRYIVROPDGGMGREYSDLEVVPITGGEHQAIDEPITAKGEHMSRALQIENDR
TSEVGKO"
/complement(4945..4948)
```

```
Query Match 31.3%; Score 31.6; DB 1; Length 23740;
Best Local Similarity 58.5%; Pred. No. 2.9e+02;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GCCTTCGGGATGATACGCCGGTGGCGGTGCAGGTCAATGGCAGCAGGACGACCGC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19925 GCCTGCTGGAATATTCGACGGGTGCACGAGCAAGCCATCGACGACTTCTGCC 19866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GCCTGCAATTCACACGTTGCCAGCCGGGTGCCCG 94
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19865 GCGCCGACACATCTTTCGATCCTCGAATCGG 19832
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: November 6, 2002, 13:24:51
Job time : 568.625 secs


```

/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGA819"
/clone_1fb="BOGA"
/notes="vector: pHD1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHD1 using BstXI linkers"
BASE COUNT      149 a      246 c      238 g      165 t
ORIGIN

```

Query MatchScore	36.2%	Score	36.6	DB	17	Length	798
Best Local Similarity	65.1%	Pred.	No. 9.5				
Matches	54	Conservative	0	Mismatches	29	Indels	0
						Gaps	0

[illegible]

```

oy      78  GTCGAGTCAAGGTGACGTCGCC 100
          || ||||| |||||
Db      756 CGCGCTACAAGGTGCTCGTCC 778

```

RESULT 2	LOCUS	DEFINITION
B1780606/c	B1780606	387 bp mRNA linear EST 23-JUL-2007
	EBES01-SQ001_G23	R embryo sac, 4-6 DPA, no treatment, cv Optic,
	EBES01	Hordeum vulgare cDNA clone EBES01-SQ001_G23 5', mRNA
		sequence.

ACCESSION	BI780606
VERSION	BI780606.2
KEYWORDS	GI:21938718
SOURCE	EST.
ORGANISM	<i>Hordeum vulgare</i> .
	<i>Hordeum vulgare</i>

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
1 (bases 1 to 387)

REFERENCE	1 (bases 1 to 387)
AUTHORS	Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cartwright, J.,
TITLE	Ramsay, L., Machry, G., Marshall, D.F.M. and Waugh, R.
JOURNAL	Development of Barley Transcriptome Resources
COMMENT	Unpublished (2001)
	On Sep 26, 2001 this sequence version replaced gi:15783498

Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: estes@scsi.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.

FEATURES	Location/Qualifiers
source	1. .387

```

/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBS01_SQ001_G33"
/clone_lib="embryo sac, 4-6 DPA, no treatment, cv Optic"

```

```
/tissue_type="embryo sac"  
/dev_stage="4-6 DPA"
```

```
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;  
Non-normalised library, directionally cloned into pSPORT1
```

Derived from embryo sacs dissected from developing grains (4-6 days post anthesis) in glasshouse grown barley

plants. developed as part of the variety transcripome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."

BASE COUNT	90 a	127 c	125 g	45 t
ORIGIN				

Query Match	35.28;	Score 35.6;	DB 13;	Length 387;
Best Local Similarity	60.28;	Pred. No. 15;		
Matches 59;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0;

Qy **4** GACGGTGTGCGCGGCGAGCCGCCTCGCAGAGTGCTGGCGATTCCTTGGGCAAA 63
 || || | ||| | ||| | ||| | ||| | ||| |
Db **291** GAGGCGTTGGCGTAGGTGCCGCGCGCATGACC GG GTGTGCGCGATTCTGCGGCGCAT 232

QY 64 GGGGGTGTTCGCCGTCAGTCACGATGGAGCTCGCCA 101
 | | | | | | | | | |
Db 231 CTTGTTACCAGCCCACTCCGTGGAGCTGCGCCAGCCA 194

RESULT 3					
BE740888/c	BE740888	979 bp	mRNA	linear	EST 15-SEP-2000
LOCUS	6015593320P1				
DEFINITION	Homo sapiens cDNA clone IMAGE:3947188 5', mRNA sequence				

ACCESSION	BE740888	
VERSION	BE740888.1	GI:10154880
KEYWORDS	EST.	
SOURCE	human	

ORGANISM

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1 (bases 1 to 979)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP

cdna Library Preparation: Ling Hong/Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL,

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MCC clone distribution information can be found at: <http://www.mccclone.com>

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 plate: LLCM808 row: h column: 05
 with small blue arrows at the 241

High quality sequence stop: 341.	
Location/Qualifiers	
1 070	

```
source      1. .979  
/organism="Homo sapiens"  
/abs-unc6="tumor-0606"
```

```

/db_xref="taxon:9606"
/clone="IMAGE:3947188"
/clone_1:b-ENTH WCC 0"

```

```
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="fusion (heterococci)"
```

```

/Lab_host="DH10B (phage-resistant)"
/Note="Organ: ovary; Vector: pOTB7; Site.1: XhoI; Site.2:
EcoRI. DNA made by standard routine. Directionally

```

ECORI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: CCGAGCTCGA. Size-selected >500bp for average

adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M Rubin (University of

the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).

BASE COUNT	ORIGIN	(Stratagene) and Superscript II RT (Life
219 a	394 c	204 g 161 t 1 others

Query match	Score	DB	Length
34.5%	34.8	12	979
33.8%	34.8	12	979
33.8%	34.8	12	979

Best Local Similarity 62.88; Pred. No. 29;
Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

10 GCTGGCGGCGA6CCGGCAGCGTCCGACAGGTCCTGGCGATCCTTGGCGGCAAAAGGGGT 69

Db 653 GGTGCGGTGGCGCGCTGGCTGCCTCACCGAGGGGGTGGGGTGGCGCAGTGGCAGTGGGGGT 594

QY 70 GTTCTGCCGTCGAGTCAAGGTGACG 95

Db 593 GGTGGCCGTGGAGCCAGGGCGCG 568

RESULT 4
AW4448711/C

LOCUS AM448711 678 bp mRNA linear EST 03-JAN-2001
 DEFINITION BR1180 BRY Triticum aestivum cDNA clone P26-1C, mRNA sequence.
 ACCESSION AM448711
 VERSION AM448711.1 GI:12019246
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Triticum.
 1 (bases 1 to 678)
 REFERENCE Clarke,B.C., Hobbs,M. and Appels,R.
 TITLE Genes active in developing wheat endosperm
 JOURNAL Unpublished (2000)
 COMMENT Contact: Bryan Clarke
 Division of Plant Industry
 C.S.I.R.O.
 GPO Box 1600, Canberra, ACT, Australia
 Tel: 61 2 6246 5054
 Fax: 61 2 6246 5000
 Email: bryan@pl.csiro.au.

FEATURES
 source
 1..678
 /organism="Triticum aestivum"
 /cultivar="Wynna"
 /db_xref="taxon:4565"
 /clone="P26-1C"
 /clone_lib="BRY"
 /cell_type="endosperm"

BASE COUNT 121 a 300 c 160 g 97 t
 ORIGIN

Query Match 33.5%; Score 33.8; DB 10; Length 678;
 Best Local Similarity 60.2%; Pred. No. 48;
 Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 3 CGACGCTGCTGCGGCGAGCGCGCTCGACAGAGTCTGCGATCTTGGCGCA 62
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 133 CGACGAGTCTGCGGCGAGCGCGCTCGACAGAGTCTGCGATCTTGGCGCA 74
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 63 AGGGGGTCTCTGCCGTGAGTCAAGTGGACG 95
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 73 CGGGGAGAGTGTGCGCGCGGGCGGTGTCG 41
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
 BE428772 401 bp mRNA linear EST 26-JUL-2000
 LOCUS MTD010.F10F990617 ITEC MTD Durum Wheat Root Library Triticum
 DEFINITION turgidum subsp. durum cDNA clone MTD010.F10, mRNA sequence.
 ACCESSION BE428772
 VERSION BE428772.1 GI:9426615
 KEYWORDS EST.
 SOURCE durum wheat.
 ORGANISM Triticum turgidum subsp. durum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Triticum.
 1 (bases 1 to 401)
 REFERENCE Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
 Hermann,R.G., Holton,T., Jacquemelin,J.M., Jia,J., Joudrier,P.,
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,T.,
 Pechioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
 Sorrells,M., Warburton,M. and Wenzel,G.
 TITLE International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 JOURNAL Unpublished (2000)
 COMMENT Contact: Joudrier P
 INRA, Unite de Biochimie et Biologie Molculaire des Cereales
 2, place VIALA, 34060 Montpellier cedex 01 FRANCE
 Tel: 33 4 99 61 23 84
 Fax: 33 4 99 61 23 48

Email: joudrierensam.inra.fr
 International Triticeae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1..401
 /organism="Triticum turgidum subsp. durum"
 /cultivar="Silliana"
 /db_xref="taxon:4567"
 /clone="MTD010.F10"
 /clone_lib="ITEC MTD Durum Wheat Root Library"
 /tissue_type="root"
 /dev_stage="3-day-old seedling, water-stressed"
 /note="Vector: pSPORT1; T7 primers used. See pSPORT1
 polylinker site. 0.3-2.0 kbp average insert size."

BASE COUNT 49 a 99 c 182 g 71 t
 ORIGIN

Query Match 33.3%; Score 33.6; DB 10; Length 401;
 Best Local Similarity 66.7%; Pred. No. 49;
 Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 2 TCGACGCTGCTGCGGCGAGCGCGCTCGACAGAGTCTGCGATCTTGGCGCA 61
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 209 TCGACGCTGCTGCGGCGAGCGCGCTCGACAGAGTCTGCGATCTTGGCGCG 268
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 62 AAGGGGGTGTTC 73
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 269 GTGGTGTGCTTC 280
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
 B1471502/c 534 bp mRNA linear EST 29-NOV-2001
 LOCUS sag21c02.y1 Gm-c1080 glycine max cDNA clone GENOME SYSTEMS CLONE
 DEFINITION ID: Gm-c1080-1659 5' similar to SW:5114.HUMAN Q15459 SPLICISOME
 ASSOCIATED PROTEIN 114 ; mRNA sequence.
 ACCESSION B1471502
 VERSION B1471502.1 GI:15287611
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 1 (bases 1 to 534)
 REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Expelding,J., Coryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available through: Reggen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4365 or contact via email: ccu@reggen.com
 High quality sequence stop: 422.
 Location/Qualifiers
 1..534
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1080-1659"
 /clone_lib="Gm-c1080"
 /tissue_type="Roots of 8 day old 'Bragg' supermodulating
 mutant NT5382 seedlings"

BASE COUNT
ORIGIN

139 a 161 c 99 g 134 t 1 others

/'lab-stage-"8 days old"
/'lab_host-"DH10B"
/note="vector: pBluescript II SK+, Site_1: EcoRI, Site_2:
XhoI; The mRNA was isolated from roots of 8 day old
'Bragg' supermodulating mutant NRS382 seedlings that were
infected with Bradyrhizobium japonicum, strain USDA 110, 72
hours prior to harvest. Dr. Gary Stacey generously donated
the tissue. The roots were flash-frozen in liquid
nitrogen. Stratagene's cDNA Synthesis Kit (catalog number
200401) was used to synthesize the cDNA. First-strand
synthesis was performed with 5-methyl dCTP, hence the
ligated cDNA was hemimethylated. A modification of
Stratagene's first-strand synthesis primer was used. An
'anchor' nucleotide (V-A, C, or G) was added to the 3' end
of the primer [GAGAGAGAGAGAGAGAGAGAACTAGCTGTGAG(T)18V] to
anchor the primer at the 5' end of the poly(A) tract.
After second-strand synthesis, the cDNA ends were filled
in with cloned Pfu DNA, ligated to EcoRI adapters and
subsequently phosphorylated. The cDNA was then
precipitated and redissolved in sterile, RNase-
free water. The XhoI site within the first-strand
synthesis primer was then restricted by digestion with
XhoI from Promega (400U/ul); all XhoI sites in the cDNA
would be protected by their hemimethylated status. The
cDNA constructs were size-fractionated with a 500bp
cutoff, using Sephacryl S-500 High Resolution (Pharmacia
Biotech) in a 2-mm diameter column and a bed volume of
approximately 1ml. The column eluent was precipitated,
redissolved, and ligated into Stratagene's pBluescript II
XR predigested vector (pBluescript II SK+) vector that
has been digested with EcoRI and XhoI, and phosphorylated
by Stratagene). This library was constructed in the
laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at
Northern Arizona University."

[illegible]

FEATURES
source
1.700
Location/Qualifiers
Email: slingh@eem.agr.ca
Fax: (613) 759-1701
Tel: (613) 759-1662
06, Canada
KW Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A

```

/cClone_11b="Gm01_AAFc_ECORc_Glycine_max_cold_stressed_leaves"
/tissue_type="Leaves"
/dev_stage="15-day seedlings"
/note="Vector: Bluescript SK-/XhoI-ECOR1; site_1: ECOR1; site_2: XhoI; Plants were grown 12 days from seeds, treated at 20c for 3 days, then leaves were collected. Library was prepared using Unilap ZAP-cDNA synthesis kit / Packaged GigaPack III Gold."
BASE COUNT      194 a      135 c      179 g      180 t      12 others
ORIGIN

```

	Query Match	33.3%	Score 33.6;	DB 12;	Length 700;	
	Best Local Similarity	70.3%	Pred. No. 55;			
	Matches 45; Conservative	0;	Mismatches 19;	Indels 0;	Gaps 0;	
QY	7 GGTGCTGGCGGGAGCCGCCAGCGCCTTCGCAGAGTGTGGCGATCTTGCGCAAAAGG	66				
Dd	444 GGGACTGGGGGAGACGCTCTCTGTGTGGCACGTGTTGTCGACCTCTTTGAGGCAAAAGT	503				
QY	67 GG TG 70					
Dd	504 GG TG 507					

RESULT	8
LOCUS	BG839262
DEFINITION	BG839262 764 bp mRNA linear EST 25-MAY-2001 Gm01_10g08_F Gm01_AAPC_ECORC_Glycine_max_cold_stressed_leaves
ACCESSION	Glycine max cDNA clone Gm01_10g08, mRNA sequence. BG839262
VERSION	BG839262.1 GI:14205584
KEYWORDS	EST.
SOURCE	soybean.
ORGANISM	Glycine max Euarvota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
REFERENCE	Glycine. 1 (bases 1 to 764)
AUTHORS	Singh,J.A., Robert,L.S., Lu,B., Zhu,L., De Moors,A., Courroux,P., Harris,L.J., Hattori,J.I., Quellet,T., Sprott,D. and Tinker,N.A.
TITLE	Expressed Sequence Tags from Cold-Stressed Glycine max (soybean) Leaves
JOURNAL	Unpublished (2001)
COMMENT	Contact: Singh,J.A. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada KW Neethy Bldg., Central Experimental Farm, Ottawa, Ontario, K1A 0C6, Canada Tel: (613) 759-1662 Fax: (613) 759-1701 Email: singha@em.agr.ca.
FEATURES	Location/Qualifiers 1..764 /organism="Glycine max" /cultivar="Maple Arrow" /db_xref="taxon:3847" /clone="Gm01_10g08" /clone_1bp="Gm01_AAPC_ECORC_Glycine_max_cold_stressed_leaf es" /tissue_type="Leaves"

```

/dev.stage="15-day seedlings"
/note="Vector: Bluescript SK-/XhoI-EcoRI; Site_1: EcoRI;
Site_2: XhoI; Plants were grown 12 days from seeds,
treated at 20C for 3 days, then leaves were collected.
Library was prepared using Unizap ZAP-CDNA synthesis kit
/ packaged GigaPack III Gold."
BASE COUNT      210 a      152 c      203 g      186 t      13 others
ORIGIN

```

Query Match	33.3%	Score 33.6;	DB 12;	Length 764;
Best Local Similarity	70.3%	Pred. No. 56;		
Matches 45;	Conservative 0;	Mismatches 19;	Indels 0;	Gaps 0;

[illegible]

OY	67	GGTG	70
Db	507	GGTG	510

RESULT 9					
BM816836					
LOCUS	BM816836	870 bp	mRNA	linear	EST 05-MAR-2002
DEFINITION	HB105A04_SK.ab1	HB Hordeum vulgare	CDNA clone	HB105A04_SK.ab1	

ACCESSION	BM616836
VERSION	BM616836.1
KEYWORDS	GI:19152850
SOURCE	EST.
	<i>Hordeum vulgare</i> .

REFERENCE
AUTHORS

Euarystola, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
+ Triticeae; Hordeum.
1 (bases 1 to 870)
Ozturk,N.Z., Michalowski,C.B., Brazille,S., Borchert,C., Palacio,C.

TITLE Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley
JOURNAL Unpublished (2002)
COMMENT Contact: Mark A. Fredricksen

1201 W Gregory Dr, Urbana, IL 61801, USA
Tel: 2172655473
Email: bohnertlab@life.uiuc.edu.
Location/Qualifiers
I. 870

BASE COUNT	ORIGIN
195 a	214 c
246 g	209 t
6 others	

Query Match	33.3%	Score 33.6	DB 14	Length 870
Best Local Similarity	66.7%	Pred. No. 57		
Matches	48	Conservative	0	Mismatches 24
			Indels	0
			Gaps	0
QY	2	TCGACGGTCGTCGGCGGAGCGCGGACGCGCCGCGACAGAGTCTCTGCGCATCTTTCGGCA	61	
DB	86	TCGACGGTCGTCGTCGGCGGAGCGGAGGAGGTCCGCTTACACAGCGCTGGCGCG	145	

QY	62	AAGGGG	GTTC	73
Db	146	GTGGTGTG	GTTC	157

RESULT 10	LOCUS	DEFINITION
C67972/c	C67972	360 bp mRNA linear EST 22-SEP-1997
	C67972	yuji Kohara unpublished CDNA <i>Caenorhabditis elegans</i> CDNA
		clone yk308c8 5', mRNA sequence.

ORGANISM	Caenorhabditis elegans.
SOURCE	Caenorhabditis elegans
KEYWORDS	EST.
VERSION	C67972.1
ACCESSION	C67972
GI	2426902

REFERENCE
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
 1 (bases 1 to 360)
 Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
 M., Miyata, A. and Nishigaki, A.
 Expression map of the *C.elegans* genome
 Unpublished (1996)
CONTACT: Yuji Kohara

Genome Biology Ltd.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp
<http://www.nig.ac.jp/koohara/koohara.htm>

FEATURES	SOURCE	LOCATION/QUALITERS
1. .360		
/organism="Caenorhabditis elegans"		
/strain="CB1489 him-8(e1489)"		
/db_xref="taxon:6239"		
/clone="yk308c8"		
/clone_id="Yuj1 Kohara unpublished cDNA"		
/sex="hermaphrodite, male"		
/tissue_type="whole animal"		
/dev_stage="varied"		
87 a	109 c	88 t
	75 g	1 others
BASE COUNT		
ORIGIN		

	Query Match	33.1%	Score 33.4;	DB 14;	Length 360;
	Best Local Similarity	68.7%;	Pred. No. 54;		
	Matches 46;	Conservative	0;	Mismatches 21;	Indels 0;
	Gaps				
Qy	7	GGTGTGCGCGGCAGCCCTGTGCAGAGTGCTTGGCGCAAAAGG	66		
Dd	149	GSTGGGGCGGCCAGGCGCGCATACCACAGCAGGTGGGGGTCCTCCATCAATGTT	90		

QY	67	GGTGTTC	73
Db	89	GGTGTTC	83

RESULT 11	
LOCUS	Bt131405
DEFINITION	Bt131405 unpublished oligo-capped cDNA library, C. elegans L1 stage
ACCESSION	Bt131405
VERSION	Bt131405
KEYWORDS	Bt131405.1 GI:18291562
SOURCE	EST.
	Caenorhabditis elegans.

ORGANISM	Caenorhabditis elegans Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditoidea : Rhabditidae; Pelodierinae; Caenorhabditis. 1 (bases 1 to 617)
REFERENCE	Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.
TITLE	A complementary view of the C.elegans genome
JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i

Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
source

1. .617
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="unpublished oligo-capped cdna library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wrmcdna library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

BASE COUNT 190 a 110 c 172 g 143 t 2 others

ORIGIN

Query Match 33.1%; Score 33.4; DB 13; Length 617;
Best Local Similarity 68.7%; Pred. No. 60;
Matches 46; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 7 GGTCTGGCGGCGAGCGCGCTGCGAGAGTGCTGCGATCCTTGGCGCAAGGG 66
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 398 GGTGGGGCGGCGAGCGCGCATACCCAGAGAGTGCGGTGCTCCCATCATGTGT 457

QY 67 GGTGTTTC 73
||||| ||
Db 458 GGTGCTC 464

RESULT 12 691 bp mRNA linear EST 23-JAN-2002
BUI32434
LOCUS BUI32434 unpublished oligo-capped cdna library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cdna clone yk1055h12 3', mRNA sequence.
ACCESSION BUI32434
VERSION BUI32434.1 GI:18292591
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
1 (Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 691)
Kohara,Y., Shin-I.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugeno,S.

REFERENCE
AUTHORS A complementary view of the C.elegans genome
Unpublished (2002)
Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
source

1. .691
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="unpublished oligo-capped cdna library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"

/note="The AD-wrmcdna library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

BASE COUNT 206 a 129 c 190 g 166 t

ORIGIN

Query Match 33.1%; Score 33.4; DB 13; Length 691;
Best Local Similarity 68.7%; Pred. No. 61;
Matches 46; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 7 GGTCTGGCGGCGAGCGCGCTGCGAGAGTGCTGCGATCCTTGGCGCAAGGG 66
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 407 GGTGGGGCGGCGAGCGCGCATACCCAGAGAGTGCGGTGCTCCCATCATGTGT 466

QY 67 GGTGTTTC 73
||||| ||
Db 467 GGTGCTC 473

RESULT 13 623 bp mRNA linear EST 07-FEB-2002
BM491932
LOCUS BM491932
DEFINITION ppp2n.pk007.n21 Normalized Chicken Pituitary/Hypothalamus/Pineal Library (ppp2n) Gallus gallus cdna clone ppp2n.pk007.n21 5' similar to ref|NP_115872.1 (NM_032483) HTPAP protein [Homo sapiens] gb|AAK14924.1|AF212238_1 (AF212238) HTPAP [Homo sapiens], mRNA sequence.
ACCESSION BM491932
VERSION BM491932.1 GI:18612863
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 623)
Porter,T.E. and Cogburn,L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cdna library, USDA/IFARS Animal Genome Project
Unpublished (2002)
Contact: Larry A. Cogburn
University of Delaware
Towsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES
source

1. .623

/organism="Gallus gallus"
/strain="Commercial broiler chickens"
/db_xref="taxon:9031"
/clone_lib="ppp2n.pk007.n21"
/clone_lib="Normalized Chicken Pituitary/Hypothalamus/Pineal Library (ppp2n)"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (1,3,5,7,9 weeks)"
/lab_host="E. coli EMD10B"
/note="Vector: PCWVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end"

BASE COUNT 72 a 210 c 210 g 131 t

ORIGIN

Query Match 32.9%; Score 33.2; DB 13; Length 623;
Best Local Similarity 67.1%; Pred. No. 68;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 3 CGACGTTGTCGGCGGCGAGCGCTGCGAGAGTGCTGCGATCCTTGGCGCAA 62
|| ||| ||||||| || ||||| || || ||| ||||| |||||

Db	26	CGGCGGCGCTGGCGCGAGAGCGCGGTCGGGCGCGCTTGGCGCGCGTGTGTGGCGGA	85
Oy	63	AGGGGCTGT 72	
Db	86	TGGACTGCT 95	
RESULT 14			
LOCUS			
BM427451			
DEFINITION			
Bm427451 660 bp mRNA linear EST 30-JAN-2002			
pf2n.pk006.m6 Normalized Chicken Abdominal Fat Library (pf2n)			
gallus.gallus cDNA clone pf2n.pk006.m6 5' similar to g111210482			
ref NP_115872.1 HRPAP protein [Homo sapiens] g1114750587			
ref XP_009446.1 HRPAP protein [Homo sapiens]			
g1 AAK14924.1 AF212238_1 (AF212238) HRPAP [Homo sapiens], mRNA			
sequence.			
BM427451			
ACCESSION			
BM427451			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Chicken.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
Phasianinae; Gallus.			
1 (bases 1 to 660)			
Cogburn,L.A., Morgan,R. and Burnside,J.			
ESTs from Normalized Chicken fat CDNA library-USDA/IFAFs Animal			
Genome Project			
Unpublished (2002)			
Contact: Larry A. Cogburn			
University of Delaware			
Townsend Hall, Newark, DE 19717, USA			
Tel: 302-831-1335			
Fax: 302-831-2832			
Email: cogburn@udel.edu, www.chickest.udel.edu.			
FEATURES			
Source			
1..660			
/organism="Gallus gallus"			
/strain="Commercial broiler, Ottawa Research Centre,			
leghorn"			
/db_xref="taxon:9031"			
/clone_lib="Normalized Chicken Abdominal Fat Library			
(pf2n)"			
/sex="Male and Female"			
/tissue_type="Abdominal Fat"			
/dev_stage="Embryonic (dl8,dl9); post-hatch (dl,w3,w7,w9			
,w16,1yr)"			
/lab_host="E. coli EMDH10B"			
/note="Vector: pCMVSPORT6; Library made from equivalent			
pools of total RNA isolated from each developmental age			
(across strains); Single pass sequencing from 5'-end"			
BASE COUNT			
77 a 218 c 222 g 138 t 5 others			
ORIGIN			
Query Match 32.9%; Score 33.2; DB 13; Length 660;			
Best Local Similarity 67.1%; Pred.No. 68;			
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;			
Oy	3	CGACGGTCTGGCGGCGGACGCGGCTTCGACAGAGTCTGGCATCTTGGCGGCA	62
Db	34	CGGCGGCGCTGGCGGCGAGCGCGGTCGGGCGCGCTTGGCGCGGTGTGTGGCGCA	93
Oy	63	AGGGGCTGT 72	
Db	94	TGGACTGCT 103	
RESULT 15			
H40050/c			
LOCUS			
DEFINITION			
Y1440D9.r1 Soares breast 3MBst Homo sapiens CDNA clone			
IMAGE:161105 5' similar to SP:AA2184 A42184; NDM PROTEIN - ; mRNA			

	sequence.	H40050	H40050.1	GI:916102
VERSION	EST.			
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	I (bases 1 to 371)			
	Hillier,L., Clark,N., Dubque,T., Elliston,K., Hawkins,M., Holman			
	,M., Hultman,M., Kucab,T., Le,M., Lennon,G., Marra,M., Parsons,J.,			
	Rickin,L., Rohling,T., Soares,M., Tan,F., Trevasakis,E., Waterston			
	,R., Williamson,A., Wohlmann,P. and Wilson,R.			
TITLE	The Mashu-Mercx EST Project			
JOURNAL	Unpublished (1995)			
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 1038 High quality sequence stops: 339 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1038 Std Error: 0.00 Seq primer: M3Rev High quality sequence stop: 339. Location/Qualifiers 1..371 /organism="Homo sapiens" /db_xref="GDB:575215" /db_xref="taxon:9606" /clone="IMAGE:161105" /clone_lib="Soares breast 3nbHBst" /sex="Female" /dev_stage="adult" /lab_host="DH10B (ampicillin resistant)" /note="Organ: breast; Vector: pYT73D (Pharmacia) with a modified polylinker; site.1: Not I; Site-2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGCGCCGCCCTTTTTCCTTTTTTTT 3'] , (Pharmacial), digested with Not I and cloned into the Not I and Eco RI sites of a modified pYT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima BonaldO."			
BASE COUNT	83 a 129 c 94 g 57 t 8 others			
ORIGIN				
Query Match	32.5%	Score 32.8:	DB 14;	Length 371;
Best Local Similarity	60.5%;	Pred. NO. 77;		
Matches 52;	Conservative 0;	Mismatches 34;	Indels 0;	Gaps 0;
OY	12 TGGCGGCGCAGCCGCACGTCCGAGAAGTGCTGCGCATCTGCGGCAAGGAGGTGT 71			
Db	258 TTGGCGCGCAGCAGTGGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 199			
OY	72 TCTGCGGTGAGTCAGAGTGAGCGTC 97			
Db	198 NCCACTGNMGAGTGTGGGGGAAACC 173			
Search completed:	November 6, 2002,	15:52:36		
Job time :	1202.62 secs			

Query Match 30.7%; Score 31; DB 10; Length 969;
Best Local Similarity 72.7%; Pred. No. 0.84;
Matches 40; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 13 GCGGGGCGACCGGACCGCTCCGACAGGTGCTGGCATCTTGGCGCAAGGGG 67
DB 513 GCGGGGTTCCGCGCGCGATCGAGCGTGGCGGCGAGCTGCCGACCAAGGGG 567

RESULT 2

US-09-747-810-1/c
; Sequence 1, Application US/09747810
; Patent No. US20020012903a1
; GENERAL INFORMATION:

; APPLICANT: Goldman, Steven A.
; APPLICANT: Okano, Hideyuki
; TITLE OF INVENTION: A METHOD FOR ISOLATING AND PURIFYING MULTIPOTENTIAL NEURAL PROGEN
; TITLE OF INVENTION: CELLS AND MULTIPOTENTIAL NEURAL PROGENITOR CELLS
; FILE REFERENCE: 19603/3580
; CURRENT APPLICATION NUMBER: US/09/747,810
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,003
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 52216
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-747-810-1

Query Match 29.1%; Score 29.4; DB 10; Length 52216;
Best Local Similarity 66.7%; Pred. No. 4;
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 6 CGGTGCTGGGGGCGAGCCGCGACCGCTCCGACAGTGTGCGATCTTGGCGCAAGG 65
DB 52183 CGGCGGGGCGGCGCGCGCGCTCGGCGGGGCGAGTGAAGGCGGCGAAGG 52124
QY 66 GGG 68
DB 52123 GGG 52121

RESULT 3

US-09-822-849A-22
; Sequence 22, Application US/09822849A
; Patent No. US20020045170a1
; GENERAL INFORMATION:

; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakr
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-22

Query Match 28.7%; Score 29; DB 10; Length 1204;

Best Local Similarity 58.8%; Pred. No. 3.3;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 9 TGCTGGCGGGCGAGCCGCGACGCTCGACAGGTGCTGGCATCTTGGCGCAAGGGG 68
DB 531 TGATGACCGCTGAGCTTGAAGCGCTGCTGAAGCGCTGGCCACACTGGGGCGAGCAAG 590

QY 69 TGTTCTGCCGCTGAGTCAAGTGA 93
DB 591 GGTTCCTGCCCGGTGTGTACGCGGA 615

RESULT 4

US-09-964-824A-252/c
; Sequence 252, Application US/09964824A
; Patent No. US20020102531a1
; GENERAL INFORMATION:

; APPLICANT: Hottigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 252
; LENGTH: 2678
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-252

Query Match 28.7%; Score 29; DB 10; Length 2678;
Best Local Similarity 58.8%; Pred. No. 3.6;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 9 TGCTGGCGGGCGAGCCGCGACGCTCGACAGGTGCTGGCATCTTGGCGCAAGGGG 68
DB 2190 TGATGACCGCTGAGCTTGAAGCGCTGCTGAAGCGCTGGCCACACTGGGGCGAGCAAG 2131
QY 69 TGTTCTGCCGCTGAGTCAAGTGA 93
DB 2130 GGTTCCTGCCCGGTGTGTACGCGGA 2106

RESULT 5

US-09-861-893-42
; Sequence 42, Application US/09861893
; Patent No. US20020045257a1
; GENERAL INFORMATION:

; APPLICANT: Feinberg, Andrew
; APPLICANT: Strichman-Almashanu, Liara
; APPLICANT: Jiaod, Shan
; TITLE OF INVENTION: METHODS FOR ASSAYING GENE IMPRINTING AND
; TITLE OF INVENTION: METHYLATED CPG ISLANDS
; FILE REFERENCE: 01107.00128
; CURRENT APPLICATION NUMBER: US/09/861,893
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 60/206,158
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/206,161
; PRIOR FILING DATE: 2000-05-22
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Homo sapiens

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1578)
; OTHER INFORMATION: n = A,T,C or G
US-09-861-893-42
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1578;
Best Local Similarity 57.3%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
OY 4 GACCGTGTGGCGGGCGAGCCGCGCTGCAGAGTGTGCGGATCTTGGCGCAA 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 56 GNCCTGTCTCCGCGCGCCGAGCGCGACGTAAGCTCCGCCAAGACTGTGCTCT 115
```

```
OY 64 GGGGTGTTCGCGCTGCGATCGATCAAGGTGG 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 GTGGCTTTCTTCTTCTTAAGTCTGTGGAG 144
```

RESULT 6

```
US-09-784-962-1
; Sequence 1, Application US/09784962
; Patent No. US20020146431A1
```

```
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
; PRIOR FILING DATE: 1999-07-04
; PRIOR APPLICATION NUMBER: 98 08777
; PRIOR FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
```

```
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-784-962-1
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
OY 1 ATCGACGGTGTGGCGGGCGAGCCGCGCTGCAGAGTGTGCGGATCTTGGCGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 AACGGTGGCGGGGTGGACGAGCCAGGCGCGCGGAGATGTGCGCAAGATGCTGC 268
```

```
OY 61 AAAGGGGTGTTCGCTGCGCTGAGTCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 GGGGGCGGTGTCTTCTTCTCGGTACG 296
```

RESULT 7

```
US-09-784-962-2
; Sequence 2, Application US/09784962
; Patent No. US20020146431A1
```

```
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
```

```
; PRIOR FILING DATE: 1999-07-04
; PRIOR APPLICATION NUMBER: 98 08777
; PRIOR FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-784-962-2
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
OY 1 ATCGACGGTGTGGCGGGCGAGCCGCGCTGCAGAGTGTGCGGATCTTGGCGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 AACGGTGGCGGGGTGGACGAGCCAGGCGCGCGGAGATGTGCGCAAGATGCTGC 268
```

```
OY 61 AAAGGGGTGTTCGCTGCGCTGAGTCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 GGGGGCGGTGTCTTCTTCTCGGTACG 296
```

RESULT 8

```
US-09-884-514-1
; Sequence 1, Application US/09884514
; Patent No. US20020146432A1
```

```
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHARREYRE, Catherine E.
; APPLICANT: CHAPUIS, Gilles E.
; APPLICANT: NEWTONARDS, Francis Mcneilly
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; TITLE OF INVENTION: REAGENTS
; FILE REFERENCE: ALLAN
; CURRENT APPLICATION NUMBER: US/09/884,514
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: FR 9800873
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: FR 9803707
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: FR 97/12382
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
```

```
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-884-514-1
```

```
Query Match          28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
OY 1 ATCGACGGTGTGGCGGGCGAGCCGCGCTGCAGAGTGTGCGGATCTTGGCGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 AACGGTGGCGGGGTGGACGAGCCAGGCGCGCGGAGATGTGCGCAAGATGCTGC 268
```

```
OY 61 AAAGGGGTGTTCGCTGCGCTGAGTCAAG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 GGGGGCGGTGTCTTCTTCTCGGTACG 296
```

RESULT 9

```
US-09-884-514-2
; Sequence 2, Application US/09884514
```

```
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/784,962
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/347,594
```

```
US-09-884-514-2
; Sequence 2, Application US/09884514
```

Patent No. US20020146432A1
GENERAL INFORMATION:
APPLICANT: ALLAN, Gordon
APPLICANT: MEEHAN, Brian
APPLICANT: CLARK, Edward
APPLICANT: HAINES, Deborah
APPLICANT: HASSARD, Lori
APPLICANT: HARDING, John
APPLICANT: CHAREYRE, Catherine E.
APPLICANT: CHAPUIS, Gilles E.
APPLICANT: NEWMONARDS, Francis McNeill
TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
TITLE OF INVENTION: REAGENTS
FILE REFERENCE: ALLAN
CURRENT APPLICATION NUMBER: US/09/884,514
CURRENT FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: FR 9800873
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: FR 9803707
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: FR 97/12382
PRIOR FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 1767
TYPE: DNA
ORGANISM: Porcine circovirus
US-09-884-514-2

Query Match 28.5%; Score 28.8; DB 10; Length 1767;
Best Local Similarity 58.0%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 ATGAGGCTGTCGGCGGCGGCGAGCGCTCGCAGAGTCTGCGCATCTTGGCGG 60
DB 209 AACGGTGGCGGGGTGACGACGAGGCGGCGGCGGCGGAGTCTGCGCATGCTGC 268
QY 61 AAAGGGGTGTCTGCGCTCGAGTCAAG 88
DB 269 GGGGGCGTCTCTTCTTCTCGGTACG 296

RESULT 10
US-09-917-800A-1636/C
Sequence 1636, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884

PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1636
LENGTH: 2540
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_017127
US-09-917-800A-1636

Query Match 28.3%; Score 28.6; DB 10; Length 2540;
Best Local Similarity 59.0%; Pred. No. 4.7;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 ATGAGGCTGTCGGCGGCGGCGAGCGCTCGCAGAGTCTTGGCATCTTGGCGG 60
DB 479 AGCGGCGGTGATGCGCGGCGGCGAGCGGCGGCGGCGGTGTGTGCGCGGCGGCGG 420
QY 61 AAAGGGGTGTCTGCGCTCGAGTCAAG 83
DB 419 AGAGTTGGGACCGCCACCAAG 397

RESULT 11
US-09-832-320-3
Sequence 3, Application US/09832320
Patent No. US20010049834A1
GENERAL INFORMATION:
APPLICANT: Crane, Edmund H.
APPLICANT: Crane, Virginia C.
TITLE OF INVENTION: Polynucleotide and Methods of Use
FILE REFERENCE: 35718/214291
CURRENT APPLICATION NUMBER: US/09/832,320
CURRENT FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: US 60/195,801
PRIOR FILING DATE: 2000-04-10
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 612
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(612)
US-09-832-320-3

Query Match 27.9%; Score 28.2; DB 10; Length 612;
Best Local Similarity 59.3%; Pred. No. 5.2;
Matches 48; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 7 GGTGCTGGCGGCGGCGGCGAGCGCTCGCAGAGTCTTGGCATCTTGGCGGCAAGGG 66
DB 369 GGGGTGGCGGAGCTACCGGCGGCGGCGGCGGCGGCGGAGTGTGTGTGTGGCGGAGGG 428
QY 67 GGTGCTTGGCGCTCGAGTCAA 87
DB 429 GCGGTACTACACCCACGCCAA 449

RESULT 12
US-09-832-320-1
Sequence 1, Application US/09832320
Patent No. US20010049834A1
GENERAL INFORMATION:
APPLICANT: Crane, Edmund H.
APPLICANT: Crane, Virginia C.
TITLE OF INVENTION: Polynucleotide and Methods of Use
US-09-832-320-1

FILE REFERENCE: 35718/214291
CURRENT APPLICATION NUMBER: US/09/832,320
CURRENT FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: US 60/195,801
PRIOR FILING DATE: 2000-04-10
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 898
TYPE: DNA
ORGANISM: Zee mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63)...(674)
US-09-832-320-1

Query Match
Best Local Similarity 27.9%; Score 28.2; DB 10; Length 898;
Matches 48; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Db 431 GGGTGGGCGAGCTACCCGCGCCGCCGAGGTGGGTGGGTGGCGAGG 490
QY 67 GGTGTTCTGCCGTCGAGTCAA 87
Db 491 GCGGTACTACACCCAGCCAA 511

RESULT 13
US-09-864-761-12032
Sequence 12032, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12032
LENGTH: 535
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010905.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-12032

Query Match
Best Local Similarity 27.7%; Score 28; DB 10; Length 535;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Db 419 GCGGGGCTGTGTGCGCCGCTCGCAGCAGTGTGGTTCCAGTGGGACGCGGCTTGGTG 478
QY 74 TGCCGTCGATCAGGTGAGCTC 97
Db 479 GGCCTGCATCAGGAAGCCGCGC 502

RESULT 14
US-09-834-975-775
Sequence 775, Application US/09834975
Patent No. US20020110815A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Brown, Jeffrey
APPLICANT: Bolt, Andrew
APPLICANT: Van Huijfel, Christophe
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-016B
CURRENT APPLICATION NUMBER: US/09/834,975
CURRENT FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/197,538
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 1046
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 775
LENGTH: 3376
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(3376)
OTHER INFORMATION: n = A,T,C or G
US-09-834-975-775

Query Match
Best Local Similarity 27.7%; Score 28; DB 10; Length 3376;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Db 65 GCGGCGCGCGCGCAGTGGGCGCGCGGTGGCAGTGGCGCGCGGTAGCGCGGTGG 124
QY 63 AGGGGGTG 70

Db 125 ACGTGGTG 132

RESULT 15
US-09-935

US-09-935-428A-24
; Sequence 24, Application US/09935428A
; Date of Invention: 09/09/2010

; Sequence 24, Application U
; Patent No. US20020106639A1
; GENERAL INFORMATION:

; Patent No. US2002010
; GENERAL INFORMATION:
; APPLICANT: WANG, IT

GENERAL INFORMATION:
APPLICANT: WANG, LI
APPLICANT: BABIUK,

APPLICANT: BABIUK,
APPLICANT: POTTER,
APPLICANT: WANG, L

APPLICANT: POTTER,
APPLICANT: WILSON;
;

APPLICANT: WILLSON
TITLE OF INVENTION:

; TITLE OF INVENTION:
; TITLE OF INVENTION:
; TITLE OF INVENTION:; TITLE OF INVENTION: ;
; FILE REFERENCE: 900 ;
; CURRENT AND PAST :

```

; FILE REFERENCE: 900
; CURRENT APPLICATION
; CURRENT FILING DATE

```

; CURRENT APPLICATION
 ; CURRENT FILING DATE
 ; PRIOR APPLICATION NUMBER

; CURRENT FILING DATE
 ; PRIOR APPLICATION NUMBER
 ; PRIOR FILING DATE:

PRIOR FILING DATE:
PRIOR APPLICATION NO:

; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:

PRIOR FILING DATE:
PRIOR APPLICATION N

; PRIOR APPLICATION NUMBER
 ; PRIOR FILING DATE:
 ; NUMBER OF PAGES TO BE

; PRIOR FILING DATE:
 ; NUMBER OF SEQ ID NO
 ; SOFTWARE: Retort

```

; NUMBER OF SEQ ID NO
; SOFTWARE: PatentIn
: SEQ ID NO 24

```

```

; SOFTWARE: PatentCln
; SEQ ID NO 24
; LENGTH: 1343

```

```

; SEQ ID NO 24
; LENGTH: 1343
; TYPE: DNA

```

```

;      TYPE: DNA
;      ORGANISM: Porcine
;

```

ORGANISM: Porcine
US-09-935-428A-24

US-09-935-428A-24

Query Match
Post Local Client

query Match
Best Local Similarity
Matches 44: Consequence

Best Local Similitude
Matches 44; Consequence

macines 44, conse
oy 1 ATCGACGGTGCT

QY	1	ATCGACGGTGCT

Db 1205 AGCGGTGGCGGG

Db 1205 AGCGGTGGCGGG

QY 61 AAAGGGGTGT

QY	61	AAAGGGGGTGT
Dh	1265	GGGGGGGCTGT

Search completed: November 6, 2002, 20:31:29
Job time : 32.125 secs

Job time : 32.125 secs

; Sequence 1, Application US/093335409

```

: Patent No. 6121029
:
: GENERAL INFORMATION:
:
: APPLICANT: Schupp, Thomas
:
: APPLICANT: Ilgon, James
:
: APPLICANT: Molnar, Istvan
:
: APPLICANT: Zirkle, Ross
:
: APPLICANT: Cyr, Devon
:
: APPLICANT: Goerlach, Joern
:
: TITLE OF INVENTION: GENES FOR THE BIOSYNT
:
: FILE REFERENCE: 4-30582A
:
: CURRENT APPLICATION NUMBER: US/09/335,409
:
: CURRENT FILING DATE: 1999-06-17
:
: NUMBER OF SEQ ID NOS: 30
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 1
:
: LENGTH: 68750
:
: TYPE: DNA
:
: ORGANISM: Sorangium cellulosum
:
:
: US-09-335-409-1

```

Query Match 99.0%; Score 100; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

qy 2 TCACGGTCTGGCGGCGACGGCGCAGAGGTGCTGGCGATCTTGGCGCA 61
 |||||
 Db 39113 TCACGGTCTGGCGGCGACGGCGCAGAGGTGCTGGCGATCTTGGCGCA 39172

Oy 62 AAGGGGCTTCTGCCGTGAGTCAGGTGGACGTCCCA 101
 |||||
Db 39173 AAGGGGTGTTCTGCCGTGAGTCAAGTGAGCGTCCCA 39212

```

RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568, 102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1

```

Query Match	99.0%;	Score 100;	DB 4;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 2.9e-19;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 2 TCGACGGTGTGGCGGGCGAGCCGCTCGCAGAGTGTCTGGCGATCCTTTGGCGCA 61
|||||
Db 39113 TCGACGGTGTGGCGGGCGAGCCGCTCGCAGAGTGTCTGGCGATCCTTTGGCGCA 39172

QY	62	AAGGGGGTCTCTGCCGTCAGTCAAGGTGGACGTCGCCA	101
Db	39173	AAGGGGGTCTCTGCCGTCAGTCAAGGTGGACGTCGCCA	392122

RESULT 4
US-09-567-969-1

Sequence 1, Application US/09567969
Patent No. 6355457
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30562A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match	99.0%;	Score 100;	DB 4;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 2.9e-19;		
Matches 100; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

0y 2 TCACGCGTCTGGCGGGCGAGCCGGCAGCGCTCCAGAGGTCTGGCGATCTTTCGGGCA 61
 |||||
 Db 39113 TCACGCGTCTGGCGGGCGAGCCGGCAGCGCTCCAGAGGTCTGGCGATCTTTCGGGCA 39172

QY 62 AAGGGGCTTCTGCCGTCAGTCAAGGTGAGCTGCCA 101
 |||||
 Db 39173 AAGGGGCTTCTGCCGTCAGTCAAGGTGAGCTGCCA 39212

```

RESULT 5
US-09-568-480-1
: Sequence 1, Application US/09568480
: Patent No. 6355458
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,480
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: type: DNA
: ORGANISM: Sorangium cellulosum
: US-09-568-480-1

```

Query Match	99.0%;	Score 100;	DB 4;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 2.9e-19;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

0y 2 TCGACGGTGCCTGGCGGGGAGCCGGCAGCGCTGCACAGAGTGCCTGGCGATCCCTTGGCGCA 61
|||||
Db 39113 TCGACGGTGCCTGGCGGGGAGCCGGCAGCGCTGCACAGAGTGCCTGGCGATCCCTTGGCGCA 39172

Qy 62 AAGGGGGTGTCTGCGCTGAGTCAAGGTGAGCGTCGCCA 101
|||||
Db 39173 AAGGGGGTGTCTGCGCTGAGTCAAGGTGAGCGTCGCCA 39212


```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          99.0%; Score 100; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGGGGCGGACCGGCGAGCTCGCAGAGTGTCGATCCTTGCGGCA 61
Db 39113 TCGACGGTGTCTGCGGGGCGGACCGGCGAGCTCGCAGAGTGTCGATCCTTGCGGCA 39172

OY 62 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 101
Db 39173 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 39212

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          99.0%; Score 100; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGGGGCGGACCGGCGAGCTCGCAGAGTGTCGATCCTTGCGGCA 61
Db 39113 TCGACGGTGTCTGCGGGGCGGACCGGCGAGCTCGCAGAGTGTCGATCCTTGCGGCA 39172

OY 62 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 101
Db 39173 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 101
```

```
Db 39173 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 39212

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          99.0%; Score 100; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCGACGGTGTCTGCGGGGCGGACCGGCGAGCTCGCAGAGTGTCGATCCTTGCGGCA 61
Db 39113 TCGACGGTGTCTGCGGGGCGGACCGGCGAGCTCGCAGAGTGTCGATCCTTGCGGCA 39172

OY 62 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 101
Db 39173 AAGGGGGTGTCTGCGCGCTCGAGTCAAGGTGAGCGTCGCCA 39212

RESULT 9
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Shenzana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes for The Biosynthesis of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
```

FILING DATE: 09-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/258,261
 FILING DATE: 08-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Melgs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: 1506/CIP6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 541-8587
 TELEFAX: (919) 541-8689
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 49377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Sorangium cellulosum
 IMMEDIATE SOURCE:
 CLONE: p98/l, pUL3, and pYKM15
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 383..760
 OTHER INFORMATION: /product= "SorR"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSS such as eryA from Saccharopolyspora erythraea."
 OTHER INFORMATION: the reductase domains of type I PKSS such as eryA from
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 927..19874
 OTHER INFORMATION: /product= "SorA"
 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSS that are known to be involved in the synthesis of polyketide compounds."
 OTHER INFORMATION: compounds."
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 942..7115
 OTHER INFORMATION: /product= "Module 1 of SorA"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 7203..12884
 OTHER INFORMATION: /product= "Module 2 of SorA"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 13455..19616
 OTHER INFORMATION: /product= "Module 3 of SorA"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 19671..46318
 OTHER INFORMATION: /product= "SorB"
 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 19870..24556
 OTHER INFORMATION: /product= "Module 1 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 24638..30820
 OTHER INFORMATION: /product= "Module 2 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 30881..35446
 OTHER INFORMATION: /product= "Module 3 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 35528..40114
 OTHER INFORMATION: /product= "Module 4 of SorB"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 40190..46318
 OTHER INFORMATION: /product= "Module 5 of SorB"

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "Sorm"
; OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1

Query Match          52.1%; Score 52.6; DB 1; Length 49377;
Best Local Similarity 70.7%; Pred. No. 3,6e+06;
Matches 70; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY      2   TCGACGCTGTGGCGGCAGCCGCCAGCGCTTCGACAGAGTCTTGCGATCCTTGCGCA 61
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       3300 TCACGGGTGCTGGCGGCAGCCGCCAGCGCTTCGACAGTCTTCACAGCATGGACGG 3359

OY      62   AAGGGGGTTCTGCGCTCGAGTCAGAGTGAAGTGAAGCTGCC 100
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       3360 AAGAAGCTTTTGTGCTGAGTGGAGTAGAGCTTGCC 3398

RESULT 10
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Ziemann, Bryan
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; EARLIER FILING DATE: 1998-08-31
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

Query Match          44.8%; Score 45.2; DB 4; Length 33529;
Best Local Similarity 66.3%; Pred. No. 0,00038;
Matches 65; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY      2   TCGACGCTGTGGCGGCAGCCGCCAGCGCTTCGACAGAGTCTTGCGATCCTTGCGCA 61
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1843 TCGACGCTGTGGCGGTGGGTGACGTGAGAGCGGTGAGAGCGGACTATGGGAGACTGACGCA 1902

OY      62   AAGGGGGTTCTGCGCTCGAGTCAGAGTGAAGTGAAGCTGCC 99
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1903 GAAGGTGTGTCGCCGGAAGTGAACGTGACTACGC 1940

RESULT 11
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
```

RESULT 13
US-08-258-2616-6
Sequence 6
Applicant US/08258261B
Patent No 5639949
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven

```

1  APPLICANT: Ryals, John Andrew
2  APPLICANT: Gaffney, Thomas Deane
3  APPLICANT: Lam, Stephen Ting
4  APPLICANT: Hammer, Phillip E.
5  APPLICANT: Uknes, Scott Joseph
6  TITLE OF INVENTION: Genes for the synthesis of
7  TITLE OF INVENTION: antipathogenic substances
8  NUMBER OF SEQUENCES: 22
9  CORRESPONDENCE ADDRESS:
10 ADDRESSSEE: Ciba-Geigy Corporation
11 STREET: 7 Skyline Drive
12 CITY: Hawthorne
13 STATE: NY
14 COUNTRY: USA
15 ZIP: 10532
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patentln Release #1.0, Version #1.25
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/258,261B
23 FILING DATE: 08-JUN-1994
24 CLASSIFICATION: 800
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/457,205
27 FILING DATE: 01-JUN-1995
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Elmer, James Scott
30 REGISTRATION NUMBER: 36,129
31 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 919-541-8614
34 TELEFAX: 919-541-8689
35 INFORMATION FOR SEQ ID NO: 6:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 28958 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: single
40 TOPOLOGY: linear
41 MOLECULE TYPE: DNA (genomic)
42 HYPOTHEICAL: NO
43 ANTI-SENSE: NO
44 US-08-258-261B-6
45
46 Query Match 33.7%; Score 34; DB 1; Length 28958;
47 Best Local Similarity 59.2%; Pred. No. 0.46;
48 Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
49
50 QY 3 CGACGCTCTGGCGGGCCAGCGGAGCGGCTCGCAGAGTCTGGCGATCCTTGGCGCA 62
51 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
52 Db 16640 CCAGCGCTGTATCGCGGCGAGCCGCCGCCGTCGCGAGCGGCTGTATCGCTACCGCGAG 16699
53 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
54 QY 63 AGGGGGTGTCTGCCGCTGAGTCAAGGTGAGCGTCGCC 100
55 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
56 Db 16700 CGCAGGTCTTCGCCGAGAGTCGCGCTGCACTACGCC 16737
57
58 RESULT 14
59 US-08-456-837-6
60 ; Sequence 6, Application US/08456837
61 ; Patent NO. 5643774
62 ; GENERAL INFORMATION:
63 APPLICANT: Schnupp, Thomas
64 APPLICANT: Ligon, James M.
65 APPLICANT: Beck, James Joseph
66 APPLICANT: Hill, Dwight Steven
67 APPLICANT: Ryals, John Andrew
68 APPLICANT: Gaffney, Thomas Deane
69 APPLICANT: Lam, Stephen Ting
70 APPLICANT: Hammer, Phillip E.
71 APPLICANT: Uknes, Scott Joseph
72 TITLE OF INVENTION: Genes for the synthesis of

```

```

; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

Query Match          33.7%; Score 34; DB 1; Length 28958;
Best Local Similarity 59.2%; Pred. No. 0.46;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 CGACGGTGTTCGCGGCGGAGCGCGCTGCGAGAGTCTGGGATCTTGGCGCAA 62
DB 16640 CCACGCTCTATCCGCGGAGCCCGCGCGCTGCGAGTCACTCGCTACCGCAG 16699
QY 63 AGGGGTTTCCTCCGCTGAGTCAAGTGGAGCTCGCC 100
DB 16700 CGCAGTCTTCCGCCGAGAGTCCGCTGACTACGCC 16737

RESULT 15
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
```

```

; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-342-6

Query Match          33.7%; Score 34; DB 1; Length 28958;
Best Local Similarity 59.2%; Pred. No. 0.46;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 CGACGGTGTTCGCGGCGGAGCGCGCTGCGAGAGTCTGGGATCTTGGCGCAA 62
DB 16640 CCACGCTCTATCCGCGGAGCCCGCGCGCTGCGAGTCACTCGCTACCGCAG 16699
QY 63 AGGGGTTTCCTCCGCTGAGTCAAGTGGAGCTCGCC 100
DB 16700 CGCAGTCTTCCGCCGAGAGTCCGCTGACTACGCC 16737
```

Search completed: November 6, 2002, 16:26:00
Job time : 1248.62 secs

	10	38.2	37.8	77536	21	AAA14651	Nucleotide sequenc
	11	36.6	36.2	4547	21	AAA14651	Nucleotide sequenc
	12	36.6	36.2	4571	21	AAA14662	Nucleotide sequenc
	13	36.6	36.2	4767	21	AAA14667	Nucleotide sequenc
	14	36.6	36.2	4818	21	AAA14669	Nucleotide sequenc
C	15	35.6	35.2	109519	22	AA508693	Mycobacterium DNA
	16	35.6	35.2	441529	22	AA199662	Mycobacterium tube
	17	35	34.7	11916	22	AAH79279	Streptomyces averm
	18	35	34.7	12381	21	AAZ58381	Streptomyces averm
	19	35	34.7	30690	21	AAAG2301	S. avermitilis ave
	20	35	34.7	30690	22	AAH79277	Streptomyces averm
	21	35	34.7	31422	21	AAAG2302	S. avermitilis ave
	22	35	34.7	31422	22	AAH79278	Streptomyces averm
	23	35	34.7	47981	22	AAF30757	Micromonospora meg
	24	35	34.7	50937	21	AAAO9465	Streptococcus olea
	25	35	34.7	53789	19	AAV21187	Amycolatopsis medi
	26	34	33.7	28598	17	AAAT06769	Sorangium cellulos
	27	34	33.7	28958	18	AAAT89966	Sorangium cellulos
	28	34	33.7	28958	21	AAAT5299	DNA sequence of So
	29	33.6	33.3	1010	20	AAV99255	DNA encoding an ac
	30	33.6	33.3	1010	21	AAAT2521	DNA sequence of th
	31	32.8	32.5	9513	22	AAAF88337	S. spinoza DNA fra
	32	32.8	32.5	50000	22	AAAF88313	S. spinoza DNA fra
	33	32.8	32.5	50000	22	AAAF88316	S. spinoza DNA fra
	34	32.8	32.5	80161	20	AAZ21501	DNA fragment of Sa
	35	32.6	32.3	4306	22	AAE30039	Nucleotide sequenc
	36	32.4	32.1	2715	22	AAHS2047	Mycobacterium tube
	37	31.8	31.1	109519	22	AAAS08693	Micromonospora DNA
C	38	31.4	31.1	24379	18	AAAT93095	Streptomyces freno
	39	31.4	31.1	24379	19	AAAT5925	Streptomyces roseo
	40	31	30.7	969	23	AAAS4056	Pseudomonas aerugi
	41	31	30.7	15672	12	AAQ10613	Rianodin receptor
C	42	30.8	30.5	125401	23	AAAT7186	Streptomyces nous
	43	30.6	30.3	1750	22	AAAS89605	DNA encoding novel
	44	30.6	30.3	27541	22	AAAD17185	Streptomyces nous
	45	30.6	30.3	65140	22	AAAD17184	Streptomyces nous

ALIGNMENTS

```

RESULT_1
ID      AAA29349 standard; DNA: 71989 BP.
XX      AAA29349;
AC       12-SEP-2000 (first entry)
DT
XX
DE        Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
XX
KM Epothilone: polyketide synthase; epoA; epob; epoc; epod; epeD; epof;
KW EpolL; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytosstatic; ds.
XX
OS Sorangium cellulosum.
FH
FH Key Location/Qualifiers
FT CDS 3..992
FT     /tag= a
FT     /label= ORF_A
FT     /product= transposase
FT     /note= "not part of the PKS"
FT
FT CDS 989..1501
FT     /tag= b
FT     /label= ORF_B
FT     /product= transposase
FT     /note= "not part of the PKS"
FT
FT CDS 1998..6263
FT     /tag= c
FT     /label= epoxA_gene
FT     /note= "encodes the loading domain"
FT
FT misc_RNA 2031..3548

```

```
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA
FT 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA
FT 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA
FT 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS
FT 6260..10493
FT /tag= h
FT /label= epob_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA
FT 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA
FT 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA
FT 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA
FT 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA
FT 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA
FT 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA
FT 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA
FT 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA
FT 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA
FT 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA
FT 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA
FT 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA
FT 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA
FT 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA
FT 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS
FT 10639..16137
FT /tag= x

FT /label= epoc_gene
FT /note= "encodes module 2"
FT misc_RNA
FT 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA
FT 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA
FT 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA
FT 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA
FT 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS
FT 16134..37907
FT /tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT misc_RNA
FT 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA
FT 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA
FT 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA
FT 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA
FT 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA
FT 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA
FT 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA
FT 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA
FT 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA
FT 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA
FT 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA
FT 29433..30287
FT /tag= ap
FT /note= "encodes ER5"
FT misc_RNA
FT 30321..30869
FT /tag= aq
FT /note= "encodes KR5"
FT misc_RNA
FT 31077..31373
FT /tag= ar
FT /note= "encodes ACP5"
FT misc_RNA
FT 31440..32807
FT /tag= as
FT /note= "encodes KS6"
FT misc_RNA
FT 33018..34067
FT /tag= at
FT /note= "encodes AT6"
FT misc_RNA
FT 34107..34676
FT /tag= au
```

```
FT      /note= "encodes DH6"
FT      misc_RNA
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      misc_RNA
FT      36705..37256
FT      /tag= av
FT      /note= "encodes KR6"
FT      misc_RNA
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACp6"
FT      CDS
FT      37912..49308
FT      /tag= ay
FT      /label= epoc_gene
FT      /note= "encodes modules 7 and 8"
FT      misc_RNA
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      misc_RNA
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      misc_RNA
FT      41341..41922
FT      /tag= bd
FT      /note= "encodes KR7"
FT      misc_RNA
FT      42181..42423

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATCGACGCTGCTGGCGGCGAGCCGCGCTCGCAGAGGTGCTGGCATCTTGGCGG; 60
Db      33500 ATCGACGCTGCTGGCGGCGAGCCGCGCTCGCAGAGGTGCTGGCATCTTGGCGG; 33559
QY      61 AAAGGGGTGTTCTGCCGTGCATCAAGGTGAGCTGCCCA 101
Db      33560 AAAGGGGTGTTCTGCCGTGCATCAAGGTGAGCTGCCCA 33600

RESULT 2
AAZ5587
ID      AAZ5587 standard; DNA: 68750 BP.
AC      AAZ55887;
XX      10-APR-2000 (first entry)
DE      Sorangium cellulosum 68.75 kb contig.
XX      Epothione biosynthesis; type I polyketide synthase; taxol substitute;
KM      anticancer; ds.
XX      Sorangium cellulosum.
OS      Location/Qualifiers
FH      Key
FT      CDS
FT      1..1826
FT      /tag= a
FT      /partial
FT      /product= "Partial Orf 1 protein (AAV58580)"
FT      /note= "No initiation codon given in the specification"
FT      CDS
FT      complement (1900..3171)
FT      /tag= b
FT      /product= "Orf 2 protein (AAV58581)"
FT      3415..5556
FT      /tag= c
FT      /product= "Orf 3 protein (AAV58582)"
FT      CDS
FT      complement (5612..5992)
FT      /tag= d
FT      /product= "Orf 4 protein (AAV58583)"
FT      6226..6675
FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
FT      /tag= f
```

```
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      CDS
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      CDS
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      CDS
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      CDS
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      CDS
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      CDS
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      CDS
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      CDS
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      CDS
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      CDS
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      CDS
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      CDS
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      CDS
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      CDS
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      CDS
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      CDS
FT      68346..68750
FT      /partial
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
FT      WO9966028-A2.
FT      23-DEC-1999.
FT      16-JUN-1999; 99WO-EP04171.
FT      18-JUN-1998; 98US-0099504.
FT      24-SEP-1998; 98US-0101631.
FT      05-FEB-1999; 99US-0118906.
FT      (NOVS ) NOVARTIS AG.
FT      PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT      PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

XX WPI: 2000-097741/08.
 DR P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
 DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
 DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
 DR AAY58592, AAY58593, AAY58594.
 PT New isolated epothilone synthase genes, used for the recombinant
 PT production of epothilone for use in cancer therapy
 XX
 PS Claim 14; Page 87-104; 174pp; English.
 CC This sequence represents a 68.75 kb contig from Sorangium cellulosum
 CC comprising 22 open reading frames (ORFs) and includes genes encoding
 CC proteins involved in the biosynthesis of epothilones. Epothilones A and
 CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
 CC starter unit; polyketides being synthesised from two-carbon building
 CC blocks, the beta-carbon of which always carries a keto group. Each round
 CC of two-carbon addition is carried out by a complex of enzymes known as
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
 CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
 CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
 CC and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
 CC formation. EPO F (AAY58579) is an epothilone macroactone oxidase, and
 CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
 CC involved in transport. Epothilones mimic the biological activity of
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic
 CC compositions. Epothilones exhibit a much lower drop in potency against a
 CC multiply drug-resistant cell line compared with taxol, and are
 CC considerably less efficiently exported from such cells by the multidrug
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
 CC epothilones as anticancer agents, they are problematical to produce on a
 CC large scale. Epothilones are too complex for industrial scale chemical
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
 CC poor yields of epothilones. The nucleic acids of the invention may be
 CC used for the recombinant production of epothilones in a heterologous host
 CC that is more amenable to fermentation.
 XX
 SQ Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
 Query Match 99.0%; Score 100; DB 21; Length 68750;
 Best Local Similarity 100.0%; Pred. No. 2.7e-18;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 TCGACGCTCTGGCGCGCAGCGCGCAGCGCTCGCAGAGGTGCTGCGATCTTGGCGCA 61
 Db 39113 TCGACGCTCTGGCGCGCGCGCGCGCGCGCTCGCAGAGGTGCTGCGATCTTGGCGCA 39172
 Oy 62 AAGGGGCTTTCTGCCGTCGAGTCAGAGTGAGCGTGCACA 101
 Db 39173 AAGGGGCTTTCTGCCGTCGAGTCAGAGTGAGCGTGCACA 39212
 RESULT 3
 ID AAV05287 standard; DNA; 49377 BP.
 AC AAV05287;
 XX
 XX 21-MAY-1998 (first entry)
 DT
 XX
 XX The soraphen biosynthesis gene cluster from Sorangium cellulosum.
 DE
 XX
 XX Polyketide synthase; PKS; biosynthesis; soraphen; SorP; SORa; SORb;
 KM SorM; biosynthetic module; beta-ketoacyl synthase; acyltransferase;
 KM ketoreductase; beta-ketone processing domain; cytoskeletal agent;
 KM antimicrobial agent; phytopathogenic fungi; transgenic plant;
 KM biological control; ss.
 XX
 XX Sorangium cellulosum.
 OS
 XX
 XX
 XX Key Location/Qualifiers
 FH CDS 383..760
 FT

FT /*tag= a
 FT /product= SorP
 FT /note= "gene product highly homologous to the
 FT reductase domains of type I PKSs such
 FT as eryA from Saccharopolyspora erythraea"
 FT 927..19874
 FT CDS
 FT /*tag= b
 FT /product= SORa
 FT /note= "gene product is highly homologous to
 FT type I PKSs that are known to be involved
 FT in the synthesis of polyketide compounds"
 FT 942..7115
 FT /*tag= c
 FT /product= "module 1 of SORa"
 FT /note= "12884
 FT 7203..12884
 FT /*tag= d
 FT /note= "module 2 of SORa"
 FT 13455..19616
 FT /*tag= e
 FT /product= "module 3 of SORa"
 FT /note= "146318
 FT 19871..46318
 FT /*tag= f
 FT /product= "module 4 of SORa"
 FT /note= "146318
 FT 19871..46318
 FT /*tag= g
 FT /product= SorB
 FT /note= "gene product is highly homologous to
 FT type I PKS genes"
 FT 19870..24556
 FT /*tag= h
 FT /product= "module 1 of SorB"
 FT /note= "30820
 FT 24638..30820
 FT /*tag= i
 FT /product= "module 2 of SorB"
 FT /note= "35446
 FT 30881..35446
 FT /*tag= j
 FT /product= "module 3 of SorB"
 FT /note= "40114
 FT 35528..40114
 FT /*tag= k
 FT /product= "module 4 of SorB"
 FT /note= "46318
 FT 40190..46318
 FT /*tag= l
 FT /product= "module 5 of SorB"
 FT /note= "47891
 FT 46851..47891
 FT /*tag= m
 FT /product= SorM
 FT /note= "gene product is homologous to the
 FT methyltransferase from Streptomyces
 FT hygroscopicus that is involved in
 FT the synthesis of the polyketide rapamycin"
 FT US5716849-A.
 FT 10-FEB-1998.
 FT 14-DEC-1996; 96US-0764233.
 FT 24-AUG-1993; 93WO-US07954.
 FT 08-JUN-1994; 94US-0258261.
 FT 09-OCT-1996; 96US-0729214.
 FT (NOVS) NOVARTIS FINANCE CORP.
 FT Beck JD, Hill DS, Ligon JM, Neff S, Ryals JA, Schnupp T;
 FT WPI: 1998-158369/14.
 FT DNA encoding Sorangium cellulosum polyketide(s) - used for, e.g.,
 FT biosynthesis of soraphen useful as antimicrobial agent against
 FT phytopathogenic fungi
 FT Claim 2; Columns 47-90; 64pp; English.
 PS
 XX
 XX The present sequence contains a cluster of genes that encode polyketide

CC synthases (PKSs) that are involved in the synthesis of soraphens in
CC Sorangium cellulosum. The proteins encoded by the present sequence are
CC SORR, SORR, SORR and SORR. SORR and SORR contain biosynthetic modules
CC which contain a beta-ketoacyl synthase, an acyltransferase, a
CC ketoreductase and an acyl carrier protein domain, as well as a beta-ketone
CC processing domain. S. cellulosum soraphens are useful as a cytosolic
CC and antimicrobial agent active against phytopathogenic fungi.
CC Soraphen-producing transgenic plants or biological control agents can
CC also be produced, which may reduce crop losses and nutritional
CC deprivation for local populations in many parts of the world.
XX
SQ Sequence 49377 BP; 7247 A; 19522 C; 14477 G; 8131 T; 0 other;
Query Match 52.1%; Score 52.6; DB 19; Length 49377;
Best Local Similarity 70.7%; Pred. No. 2.6e-05;
Matches 70; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 2 TCAGAGGTGCTGCGGCGAGCCGCGCTCCGACAGGTGCTGCGGATCTTGGCGCA 61
Db 3300 TCACGGGTGCTGCGGCGAGCCGCGCTCCGACAGGTGCTGCGGATCTGAGCGCG 3359
QY 62 AAGGGGTGCTTCTCCCTCGAGTCAAGGTGAGCTGCC 100
Db 3360 AAGACGCTCTTCTGCTGCGAGTGGCGATGACGTTGCC 3398
RESULT 4
AAF90036
ID AAF90036 standard; DNA; 5292 BP.
XX
AC AAF90036;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
KM Metabolic pathway operon; polyketide; polyketide antibiotic;
KM type I polyketide synthase; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..5292
FT /tag= a
FT /product= "type I polyketide synthase"
XX
FT WO200140497-A2.
XX
PN 07-JUN-2001.
XX
PD 27-NOV-2000; 2000MO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuppille K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
DR P-PDB: AAB83973.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics
XX
PS Claim 35; Page 305-307; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in

CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthase) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a26g1, and encode type I polyketide synthases.
XX
SQ Sequence 5292 BP; 905 A; 1715 C; 1725 G; 947 T; 0 other;
Query Match 47.5%; Score 48; DB 22; Length 5292;
Best Local Similarity 68.8%; Pred. No. 0.00038;
Matches 66; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 5 ACGGTGCTGCGGCGAGCCGCGCTCCGACAGGTGCTGCGGATCTTGGCGCAAG 64
Db 2134 ACCGTCCTGCTGCGGCGAGCGAGCGACTGTGCGGCAACTGCTGCGGAACTGAGCGCAA 2193
QY 65 GGGGTGCTTCTCCCTCGAGTCAAGGTGAGCTGCC 100
Db 2194 GACGCTTCTGCTCCGCTCGCGGTGAAGTGACATGCGC 2229
RESULT 5
AAF90033
ID AAF90033 standard; DNA; 34071 BP.
XX
AC AAF90033;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KM Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
KM Synthetic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..34071
FT /tag= a
FT /product= "type I polyketide synthase"
XX
FT WO200140497-A2.
XX
PN 07-JUN-2001.
XX
PD 27-NOV-2000; 2000MO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerinneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuppille K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
XX
XX Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX passing nucleic acid-enriched fractions through an anion exchange
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthase) or polypeptides, particularly for production of therapeutic


```
FT /note= "fkbc gene"  
FT complement (13452..13662)  
FT /*tag= n  
FT /note= "ACP6"  
FT complement (13761..14394)  
FT /*tag= o  
FT /note= "KR6"  
FT complement (14517..15294)  
FT /*tag= p  
FT /note= "ER6"  
FT complement (15438..16587)  
FT /*tag= q  
FT /note= "dehydratase domain (DH) 6"  
FT complement (16587..17820)  
FT /*tag= r  
FT /note= "acyltransferase domain (AT) 6"  
FT complement (17820..19053)  
FT /*tag= s  
FT /note= "KS6"  
FT complement (19116..19326)  
FT /*tag= t  
FT /note= "ACP5"  
FT complement (19464..20097)  
FT /*tag= u  
FT /note= "KR5"  
FT complement (20241..21420)  
FT /*tag= v  
FT /note= "DH5"  
FT complement (21420..22653)  
FT /*tag= w  
FT /note= "AT5"  
FT complement (22653..23892)  
FT /*tag= x  
FT /note= "KS5"  
FT complement (23992..46573)  
FT /*tag= y  
FT /note= "fkbc gene"  
FT complement (24163..24373)  
FT /*tag= z  
FT /note= "ACP4"  
FT complement (24997..26146)  
FT /*tag= aa  
FT /note= "DH4 (inactive)"  
FT complement (26146..27430)  
FT /*tag= ab  
FT /note= "AT4"  
FT complement (27430..28684)  
FT /*tag= ac  
FT /note= "KS4"  
FT complement (28750..28960)  
FT /*tag= ad  
FT /note= "ACP3"  
FT complement (29092..29740)  
FT /*tag= ae  
FT /note= "KR3"  
FT complement (29869..31018)  
FT /*tag= af  
FT /note= "DH3 (inactive)"  
FT complement (31018..32185)  
FT /*tag= ag  
FT /note= "AT3"  
FT complement (32185..33439)  
FT /*tag= ah  
FT /note= "KS3"  
FT complement (33505..33715)  
FT /*tag= ai  
FT /note= "ACP2"  
FT complement (33823..34480)  
FT /*tag= aj  
FT /note= "KR2"  
FT complement (34606..35749)  
FT /*tag= ak  
FT complement (28750..28960)  
FT misc_feature
```

```
FT /*tag= al  
FT /note= "DH2 (inactive)"  
FT complement (35749..37144)  
FT /*tag= am  
FT /note= "AT2"  
FT complement (37145..38296)  
FT /*tag= an  
FT /note= "KS2"  
FT complement (38371..38581)  
FT /*tag= ao  
FT /note= "ACP1"  
FT complement (38677..39307)  
FT /*tag= ap  
FT /note= "KR1"  
FT complement (39442..40609)  
FT /*tag= aq  
FT /note= "DH1"  
FT complement (40609..41842)  
FT /*tag= ar  
FT /note= "AT1"  
FT complement (41842..43093)  
FT /*tag= as  
FT /note= "KS of extender module 1 (KS1)"  
FT complement (43144..43660)  
FT /*tag= at  
FT /note= "ACP of loading domain"  
FT complement (43777..44629)  
FT /*tag= au  
FT /note= "ER of loading domain"  
FT complement (44974..46573)  
FT /*tag= av  
FT /note= "CoA ligase of loading domain"  
FT 46754..47788  
FT /*tag= aw  
FT /note= "fkbc gene"  
FT 47785..52272  
FT /*tag= ax  
FT /note= "fkbc gene"  
FT 52275..71465  
FT /*tag= ay  
FT /note= "fkba gene"  
FT 52362..53576  
FT /*tag= az  
FT /note= "KS7"  
FT 53577..54716  
FT /*tag= bb  
FT /note= "AT7"  
FT 54717..55871  
FT /*tag= ba  
FT /note= "DH7"  
FT 56019..56819  
FT /*tag= bc  
FT /note= "ER7"  
FT 56943..57575  
FT /*tag= bd  
FT /note= "KR7"  
FT 57710..57920  
FT /*tag= be  
FT /note= "ACP7"  
FT 57990..59243  
FT /*tag= bf  
FT /note= "KS6"  
FT 59244..60398  
FT /*tag= bg  
FT /note= "AT6"  
FT 60399..61412  
FT /*tag= bh  
FT /note= "DH8 (inactive)"  
FT misc_feature
```

Query Match 47.3%; Score 47.8; DB 21; Length 77536;
Best Local Similarity 67.7%; Pred. No. 0.00056;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2 TCACGCGTCTGGCGGCGAGCCGCTGCAGAGTCTGGCGATCTTGGCGCA 61
 DB 41137 TCACGCGTCTGGCGGCGAGCCGCTGCAGAGTCTGGCGATCTTGGCGCA 41078
 QY 62 AAGGGGTGTCTTCGCGTGCAGTCAAGTGTGACGTGCC 100
 DB 41077 GATGGGTGTCTTCGCGTGCATTCGCGTGCAGTCAAGC 41039

RESULT 8
 AAS17367
 ID AAS17367 standard; DNA: 33529 BP.
 AC AAS17367;
 XX
 DT 12-MAR-2002 (first entry)
 DE DNA sequence of S. cellulosum polyketide synthase cosmid, PKOS28-26.
 XX
 KM Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
 KM acyl transferase domain; dehydratase domain; ketoreductase domain;
 KM acyl carrier protein domain; PKOS28-26; ds.
 XX
 OS Sorangium cellulosum.
 XX
 PN US6280999-B1.
 XX
 PD 28-AUG-2001.
 XX
 PF 31-AUG-1998; 98US-0144085.
 XX
 PR 22-JAN-1998; 98US-0010809.
 XX
 PA (KOSA-) KOSAN BIOSCIENCE.
 XX
 PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
 DR WPI: 2001-606536/59.
 XX
 PT Novel purified, isolated DNA molecule from Sorangium cellulosum having
 PT polyketide open reading frame encoding modules with one or more domains
 PT such as ketosynthase, acyl transferase and acyl carrier protein domains
 PT

Claim 4; Fig 1; 72pp; English.

The present invention relates to the isolation of novel Sorangium cellulosum polyketide synthases (PKS), and the polynucleotide sequences encoding them. The polyketide synthases include catalytic domains such as ketosynthase domain, acyl transferase domain, dehydratase domain, ketoreductase domain and acyl carrier protein domain. A host cell comprising a PKS ORF (open reading frame) which encodes one or more more PKS domains is useful for producing polyketide synthases from which polyketides can be produced. The host cells are useful for constructing a library, where each individual colony of the library represents a colony with the ability to produce a particular PKS synthase and ultimately a particular polyketide. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity. Colonies in the library are also induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of candidate polyketides which can be screened for binding to desired targets such as receptors, signalling proteins, etc. The present sequence represents the DNA sequence of cosmid PKOS28-26 which encodes one or more domains of S. cellulosum PKS.

Note: The present sequence is said to encode the functional domains of S. cellulosum PKS which correspond to domains or domain subsets of the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).

Sequence 33529 BP: 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

Query Match 44.8%; Score 45.2; DB 23; Length 33529;
 Best Local Similarity 66.3%; Pred. No. 0.0027;

Matches 65; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 TCACGCGTCTGGCGGCGAGCCGCTGCAGAGTCTGGCGATCTTGGCGCA 61
 DB 1843 TCACGCGTCTGGCGGCGAGCCGCTGCAGAGTCTGGCGATCTTGGCGCA 1902
 QY 62 AAGGGGTGTCTTCGCGTGCAGTCAAGTGTGACGTGCC 99
 DB 1903 GAAGGTGTCTTCGCGGGAAGTGCAGTGCAGTCAAGC 1940

RESULT 9
 AAF90037
 ID AAF90037 standard; DNA: 6462 BP.
 XX
 AC AAF90037;
 XX
 DT 06-AUG-2001 (first entry)
 DE Nucleotide sequence of a type I polyketide synthase.
 XX
 KM Metabolic pathway operon; polyketide; polyketide antibiotic;
 KM type I polyketide synthase; ss.
 XX
 OS unidentified.
 XX
 FH Key Location/Qualifiers
 FT 1..6462
 FT CDS /tag= a
 FT /product= "type I polyketide synthase"
 FT /trans_except= "(1..3, aa: Met)"
 XX
 PN WO200140497-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 27-NOV-2000; 2000WO-FR0311.
 XX
 PR 29-NOV-1999; 99FR-0015032.
 PR 07-JUN-2000; 2000US-0209800.
 XX
 PA (AVENTIS PHARMA SA.
 XX
 PI Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
 PI Cappellano C, Francou F, Raynal A, Bail M, Sezonov G, Tiphille K;
 PI Frostegard A;
 DR WPI: 2001-374849/39.
 DR P-PSDB: AAB83974.
 XX
 PT Collection of nucleic acids from environmental samples, useful for
 PT identifying e.g. genes encoding polyketide synthases and derived
 PT antibiotics
 PT

Claim 35; Page 307-309; 356pp; French.

The specification describes a method for the preparation of a collection of nucleic acids from organisms in a soil sample. The method comprises milling a dried sample to produce microparticles; suspending these in liquid buffer; extraction of nucleic acids from the microparticle; passing nucleic acid-containing solution through a molecular sieve; passing nucleic acid-enriched fractions through an anion exchange chromatography material; and recovering fractions containing purified nucleic acids. The nucleic acids are sources for sequences that encode either operons involved in a metabolic pathway (specifically polyketide synthesis) or polypeptides, particularly for production of therapeutic or agricultural compounds, especially polyketide antibiotics. AAF90034-39 a26g1, and encode type I polyketide synthases.

Sequence 6462 BP: 1073 A; 2067 C; 2135 G; 1187 T; 0 other;

Query Match 37.8%; Score 38.2; DB 22; Length 6462;

Best Local Similarity 61.6%; Pred. No. 0.19; Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;			
QY	2	TCGACGGTGTGCGGCGAGCCGCGACCGCTCCGACAGAGTGTGGCGATCTTCCGCGCA	61
DB	2299	TCCAACCGTCATCTCCGGTGAAGTGAAGACCCCTGGCGAGGTCTGCGACGCTGGAGCGG	2358
QY	62	AAGGCGGTGTCTCCGCTCGAGTCAGGTGAGCTGCC	100
DB	2359	CGAGGCGTGTCTGCGCGCGGTGAAGTGAAGTCTGCGC	2397
RESULT 10			
ID	AAA14651	standard; DNA: 77536 BP.	
XX	AC	AAA14651;	
XX	DT	08-AUG-2000 (first entry)	
XX	DE	Nucleotide sequence of the FK-520 biosynthetic gene cluster.	
KW	FK-520;	polyketide synthase; PKS; gene cluster; immunosuppressant;	
KW	Streptomyces hygroscopicus var. ascomyceticus;	immunophilin;	
KW	FK-506 binding protein; polyketide compound; transplant rejection;	graft-versus-host disease; uveitis; alopecia universalis;	
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;	multiple sclerosis; primary biliary cirrhosis; scleroderma;	
KW	neurite outgrowth; nerve regrowth; Parkinson's disease;	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;	
KW	peripheral neuropathy; ss.		
XX	OS	Streptomyces hygroscopicus.	
XX	PH	Location/Qualifiers	
FT	Key	complement (412..1836)	
FT	CDS	/*tag- a	
FT	FT	/note= "fkbw gene"	
FT	FT	complement (2020..3579)	
FT	CDS	/*tag- b	
FT	FT	/note= "fkbv gene"	
FT	FT	3969..4496	
FT	FT	/*tag- c	
FT	FT	/note= "fkbR2 gene"	
FT	CDS	complement (4595..5488)	
FT	FT	/*tag- d	
FT	FT	/note= "fkbR1 gene"	
FT	FT	5601..6818	
FT	CDS	/*tag- e	
FT	FT	/note= "fkbE gene"	
FT	FT	6808..8052	
FT	FT	/*tag- f	
FT	FT	/note= "fkbF gene"	
FT	CDS	8156..8824	
FT	FT	/*tag- g	
FT	FT	/note= "fkbG gene"	
FT	CDS	complement (9122..9883)	
FT	FT	/*tag- h	
FT	FT	/note= "fkbH gene"	
FT	FT	complement (9894..10994)	
FT	CDS	/*tag- i	
FT	FT	/note= "fkbI gene"	
FT	FT	complement (10987..11247)	
FT	CDS	/*tag- j	
FT	FT	/note= "fkbJ gene"	
FT	FT	complement (11244..12092)	
FT	CDS	/*tag- k	
FT	FT	/note= "fkbK gene"	
FT	FT	complement (12113..13150)	
FT	CDS	/*tag- l	
FT	FT	/note= "fkbL gene"	
FT	FT	complement (13212..23988)	
FT	CDS	/*tag- m	

FT	FT	/note= "fkbG gene"
FT	FT	complement (13452..13662)
FT	FT	/*tag- n
FT	FT	/note= "ACPe"
FT	FT	complement (13761..14394)
FT	FT	/*tag- o
FT	FT	/note= "KR6"
FT	FT	complement (14517..15294)
FT	FT	/*tag- p
FT	FT	/note= "ER6"
FT	FT	complement (15438..16587)
FT	FT	/*tag- q
FT	FT	/note= "dehydratase domain (DH) 6"
FT	FT	complement (16587..17820)
FT	FT	/*tag- r
FT	FT	/note= "acylttransferase domain (ATr) 6"
FT	FT	complement (17820..19053)
FT	FT	/*tag- s
FT	FT	/note= "KS6"
FT	FT	complement (19116..19326)
FT	FT	/*tag- t
FT	FT	/note= "ACP5"
FT	FT	complement (19464..20097)
FT	FT	/*tag- u
FT	FT	/note= "KR5"
FT	FT	complement (20241..21420)
FT	FT	/*tag- v
FT	FT	/note= "DH5"
FT	FT	complement (21420..22653)
FT	FT	/*tag- w
FT	FT	/note= "AT5"
FT	FT	complement (22653..23892)
FT	FT	/*tag- x
FT	FT	/note= "KS5"
FT	FT	complement (23992..46573)
FT	FT	/*tag- y
FT	FT	/note= "fkbB gene"
FT	FT	complement (24163..24373)
FT	FT	/*tag- z
FT	FT	/note= "ACP4"
FT	FT	complement (24997..26146)
FT	FT	/*tag- aa
FT	FT	/note= "DH4 (inactive)"
FT	FT	complement (26146..27430)
FT	FT	/*tag- ab
FT	FT	/note= "AT4"
FT	FT	complement (27430..28684)
FT	FT	/*tag- ac
FT	FT	/note= "KS4"
FT	FT	complement (28750..28960)
FT	FT	/*tag- ad
FT	FT	/note= "ACP3"
FT	FT	complement (29092..29740)
FT	FT	/*tag- ae
FT	FT	/note= "KR3"
FT	FT	complement (29869..31018)
FT	FT	/*tag- af
FT	FT	/note= "DH3 (inactive)"
FT	FT	complement (31018..32185)
FT	FT	/*tag- ag
FT	FT	/note= "AT3"
FT	FT	complement (32185..33439)
FT	FT	/*tag- ah
FT	FT	/note= "KS3"
FT	FT	complement (33505..33715)
FT	FT	/*tag- ai
FT	FT	/note= "ACP2"
FT	FT	complement (33823..34480)
FT	FT	/*tag- aj
FT	FT	/note= "KR2"
FT	FT	complement (34606..35749)
FT	FT	/*tag- ak
FT	FT	complement (28750..28960)
FT	FT	misc_feature

FT	FT	misc_feature	/note= "fkbC gene"
FT	FT	FT	complement (13452..13662)
FT	FT	FT	/*tag- n
FT	FT	FT	/note= "ACP6"
FT	FT	FT	complement (13761..14394)
FT	FT	FT	/*tag- o
FT	FT	FT	/note= "KR6"
FT	FT	FT	complement (14517..15294)
FT	FT	FT	/*tag- p
FT	FT	FT	/note= "ER6"
FT	FT	FT	complement (15438..16587)
FT	FT	FT	/*tag- q
FT	FT	FT	/note= "dehydratase domain (DH) 6"
FT	FT	FT	complement (16587..17820)
FT	FT	FT	/*tag- r
FT	FT	FT	/note= "acyltransferase domain (AT) 6"
FT	FT	FT	complement (17820..19053)
FT	FT	FT	/*tag- s
FT	FT	FT	/note= "KS6"
FT	FT	FT	complement (19116..19326)
FT	FT	FT	/*tag- t
FT	FT	FT	/note= "ACP5"
FT	FT	FT	complement (19464..20097)
FT	FT	FT	/*tag- u
FT	FT	FT	/note= "KR5"
FT	FT	FT	complement (20241..21420)
FT	FT	FT	/*tag- v
FT	FT	FT	/note= "DH5"
FT	FT	FT	complement (21420..22653)
FT	FT	FT	/*tag- w
FT	FT	FT	/note= "AT5"
FT	FT	FT	complement (22653..23892)
FT	FT	FT	/*tag- x
FT	FT	FT	/note= "KS5"
FT	FT	FT	complement (23992..46573)
FT	FT	FT	/*tag- y
FT	FT	FT	/note= "fkbB gene"
FT	FT	FT	complement (24163..24373)
FT	FT	FT	/*tag- z
FT	FT	FT	/note= "ACP4"
FT	FT	FT	complement (24997..26146)
FT	FT	FT	/*tag- aa
FT	FT	FT	/note= "DH4 (inactive)"
FT	FT	FT	complement (26146..27430)
FT	FT	FT	/*tag- ab
FT	FT	FT	/note= "AT4"
FT	FT	FT	complement (27430..28684)
FT	FT	FT	/*tag- ac
FT	FT	FT	/note= "KS4"
FT	FT	FT	complement (28750..28960)
FT	FT	FT	/*tag- ad
FT	FT	FT	/note= "ACP3"
FT	FT	FT	complement (29092..29740)
FT	FT	FT	/*tag- ae
FT	FT	FT	/note= "KR3"
FT	FT	FT	complement (29869..31018)
FT	FT	FT	/*tag- af
FT	FT	FT	/note= "DH3 (inactive)"
FT	FT	FT	complement (31018..32185)
FT	FT	FT	/*tag- ag
FT	FT	FT	/note= "AT3"
FT	FT	FT	complement (32185..33439)
FT	FT	FT	/*tag- ah
FT	FT	FT	/note= "KS3"
FT	FT	FT	complement (33505..33715)
FT	FT	FT	/*tag- ai
FT	FT	FT	/note= "ACP2"
FT	FT	FT	complement (33823..34480)
FT	FT	FT	/*tag- aj
FT	FT	FT	/note= "KR2"
FT	FT	FT	complement (34606..35749)
FT	FT	FT	/*tag- ak
FT	FT	FT	complement (28750..28960)

```

FT      /tag= a1
FT      /note= "DH2 (inactive)"
FT      /complement (35749..37144)
FT      /tag= am
FT      /note= "AT2"
FT      /complement (37145..38296)
FT      /tag= an
FT      /note= "KS2"
FT      /complement (38371..38581)
FT      /tag= ao
FT      /note= "ACPI"
FT      /complement (38677..39307)
FT      /tag= ap
FT      /note= "KRI"
FT      /complement (39442..40609)
FT      /tag= aq
FT      /note= "DH1"
FT      /complement (40609..41842)
FT      /tag= ar
FT      /note= "AT1"
FT      /complement (41842..43093)
FT      /tag= as
FT      /note= "KS of extender module 1 (KSI)"
FT      /complement (43144..43860)
FT      /tag= at
FT      /note= "ACP of loading domain"
FT      /complement (43777..44629)
FT      /tag= au
FT      /note= "ER of loading domain"
FT      /complement (44974..46573)
FT      /tag= av
FT      /note= "COA ligase of loading domain"
FT      /tag= aw
FT      /note= "fkbD gene"
FT      /tag= ax
FT      /note= "fkbP gene"
FT      /tag= ay
FT      /note= "fkbA gene"
FT      /tag= az
FT      /note= "KS7"
FT      /tag= ba
FT      /note= "DH7"
FT      /tag= bc
FT      /note= "ER7"
FT      /tag= bd
FT      /tag= be
FT      /note= "ACPI"
FT      /tag= bf
FT      /note= "KS8"
FT      /tag= bg
FT      /note= "AT8"
FT      /tag= bh
FT      /note= "DH8 (inactive)"

```

Query Match 37.8%; Score 38.2; DB 21; Length 77536;
 Best Local Similarity 61.6%; Pred. No. 0.24; Indels 0; Gaps 0;
 Matches 61; Conservative 0; Mismatches 38;

```

QY      2 TCGACGGTGTGGCGGCGAGCCGCGCTTCGACAGAGTGTGGCATCTTGGCGCA 61
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      64557 TCGACAGTGTGTGGCGGCGCGGTGGAGAGTGTGTGACGCGTATGAGACC 64616
QY      62 AAGGGGTGTTCGCCGTGACGTCAAGTGTGAGTGTGCC 100
      || |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      64617 GAAGGCGTGGAGTGTGTGTATGCGCGTACTGACCC 64655

RESULT 11
AA14664
ID      AA14664 standard; DNA: 4547 BP.
AC      AA14664;
DT      08-AUG-2000 (first entry)
XX      Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX      FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
XX      immunophilin; FK-506 binding protein; polyketide compound; uveitis;
XX      transplant rejection; graft-versus-host disease; alopecia universalis;
XX      autoimmune chronic active hepatitis; inflammatory bowel disease;
XX      multiple sclerosis; primary biliary cirrhosis; scleroderma;
XX      neurite outgrowth; nerve regrowth; Parkinson's disease;
XX      Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
XX      peripheral neuropathy; ss.
XX      Synthetic.
OS      Streptomyces hygroscopicus.
FH      Key Location/Qualifiers
FT      CDS 9..4535
FT      /tag= a
FT      /note= "no termination codon given"
XX      WO200020601-A2.
XX      13-APR-2000.
XX      01-OCT-1999; 99WO-US22886.
XX      02-OCT-1998; 98US-0102748.
XX      11-MAR-1999; 99US-0123810.
XX      17-JUN-1999; 99US-0139650.
XX      (KOSA-) KOSAN BIOSCIENCES INC.
XX      Reeves C, Chu D, Khosla C, Santl D, Wu K;
PI      WPI; 2000-317716/27.
DR      P-PSDB; AAY84728.
XX      New isolated polyketide synthase nucleic acid and polyketide compounds,
XX      useful for treating e.g. transplant rejection, uveitis, multiple
XX      sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
XX      peripheral neuropathy -
PS      Example 1; Page 85-88; 126pp; English.
XX      The present sequence represents an NheI-XhoII fragment that encodes
XX      module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
XX      endogenous acyltransferase (AT) domain replaced by the AT domain of
XX      module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
XX      FK-506 is a potent immunosuppressant, and acts through initial formation
XX      of an intermediate complex with protein immunophilins known as FK-506
XX      binding proteins. The nucleic acids are used for producing polyketide
XX      compounds. The polyketide compounds can be used as immunosuppressants to
XX      prevent or treat transplant rejection, graft-versus-host disease or
XX      uveitis. They can also be used for treating e.g. alopecia universalis,
XX      autoimmune chronic active hepatitis, inflammatory bowel disease,
XX      multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
XX      also have neurotrophic activity and can be used to promote neurite

```

XX
PS Example 1: Page 70-83: 126mm: End14ch

PR 11-MAR-1999; 99US-0123810.
PR 17-TUN-1999; 99US-0120650

XX
PS Example 1: Page 70-83: 126mm: End14ch

```

XX (KOSA-) KOSAN BIOSCIENCES INC.
PA
XX
XX Reeves C, Chu D, Khosla C, Santl D, Wu K;
PI
XX
XX WPI: 2000-317716/27.
DR
XX P-PSDB: AAY84731.
XX
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -
XX
XX Example 2; Page 96-99; 126pp; English.
XX
XX The present sequence represents module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
CC and acts through initial formation of an intermediate complex with
CC protein immunophilins known as FK-506 binding proteins. The nucleic
CC acids are used for producing polyketide compounds. The polyketide
CC compounds can be used as immunosuppressants to prevent or treat
CC transplant rejection, graft-versus-host disease or uveitis. They can
CC also be used for treating e.g. alopecia universalis, autoimmune
CC chronic active hepatitis, inflammatory bowel disease, multiple
CC sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
XX Sequence 4767 BP; 731 A; 1945 C; 1468 G; 623 T; 0 other;
SO
Query Match 36.2%; Score 36.6; DB 21; Length 4767;
Best Local Similarity 60.6%; Pred. No. 0.5;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 2 TCGACGCTGCTGGCGGCGACCGCGCTGCGAGAGTGTGCGATCCTTGCGGCA 61
DB 2454 TCCACCGTGATCGCGGCGACCCCGGAAGCGGTGACCATGCTCTACCCGCTATGAGCA 2513
OY 62 AAGGGGTGTTCTGCGCGAGTCAAGTGAAGTGC 100
DB 2514 CAAGGGGTGCGGTCGCGCGCATCACCGTCGACTATGCC 2552

```

```

RESULT 14
AAAL4669
ID AAAL4669 standard; DNA: 4818 BP.
XX
XX AAAL4669;
XX
XX 08-AUG-2000 (first entry)
XX
XX Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
DE
XX
XX FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
XX Synthetic.
OS Streptomyces hygroscopicus.
OS
XX Key Location/Qualifiers
FH

```

```

FT CDS 3..4817
FT FT /*tag=a
FT FT /note="no termination codon given"
XX
XX WO200020601-A2.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-US22886.
XX
XX 02-OCT-1998; 98US-0102748.
XX 11-MAR-1998; 99US-0122810.
XX 17-JUN-1999; 99US-0139650.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Reeves C, Chu D, Khosla C, Santl D, Wu K;
XX
XX WPI: 2000-317716/27.
XX P-PSDB: AAY84733.
XX
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -
XX
XX Example 2; Page 102-105; 126pp; English.
XX
XX The present sequence represents module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
CC and acts through initial formation of an intermediate complex with
CC protein immunophilins known as FK-506 binding proteins. The nucleic
CC acids are used for producing polyketide compounds. The polyketide
CC compounds can be used as immunosuppressants to prevent or treat
CC transplant rejection, graft-versus-host disease or uveitis. They can
CC also be used for treating e.g. alopecia universalis, autoimmune
CC chronic active hepatitis, inflammatory bowel disease, multiple
CC sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
XX Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 other;
SO
Query Match 36.2%; Score 36.6; DB 21; Length 4818;
Best Local Similarity 60.6%; Pred. No. 0.5;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 2 TCGACGCTGCTGGCGGCGACCGCGCTGCGAGAGTGTGCGATCCTTGCGGCA 61
DB 2505 TCCACCGTGATCGCGGCGACCCCGGAAGCGGTGACCATGCTCTACCCGCTATGAGCA 2564
OY 62 AAGGGGTGTTCTGCGCGAGTCAAGTGAAGTGC 100
DB 2565 CAAGGGGTGCGGTCGCGCGCATCACCGTCGACTATGCC 2603

```

```

RESULT 15
AAS08693/c
ID AAS08693 standard; DNA: 109519 BP.
XX
XX AAS08693;
XX
XX 26-SEP-2001 (first entry)
XX
XX Micromonospora DNA encoding biosynthetic enzymes for Evernimycin.
DE
XX

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_33500_33600

Perfect score: 101

Sequence: 1 atcgacggtcgtcgcgcgcg.....agtcaggtgacgtcgcca 101

Scoring table:

IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_in:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1	AF217189
2	101	100.0	71989	6	AF172664
3	100	99.0	68750	1	AF210843
4	100	99.0	68750	6	AR193029
5	100	99.0	68750	6	AR199551
6	100	99.0	68750	6	AR199559
7	100	99.0	68750	6	AR199567
8	100	99.0	68750	6	AR201097
9	100	99.0	68750	6	AR208671
10	72.8	72.1	10910	1	AX024383
11	72.8	72.1	10910	6	AX024276
12	58	57.4	66808	1	SAU421825
13	52.6	52.1	49377	6	188042
14	52.6	52.1	67523	1	SCU24241
15	48.4	47.9	49736	1	AF319998
16	48	47.5	5292	6	AX153793
17	48	47.5	5292	6	AX153790
18	48	47.5	5292	6	AX153789
19	47.8	47.3	77534	1	AF235504
20	46.2	45.7	42603	1	AF188287
21	45.2	44.8	33529	6	AR166425
22	42.4	42.0	28732	1	AF082100
23	41.4	41.0	107379	1	SHGCP1R
24	39.8	39.4	104326	1	AB070940
25	38.2	37.8	6462	6	AX153794
26	38.2	37.8	18791	1	SPFKBD
27	38.2	37.8	77534	1	AF235504
28	37.6	37.2	10029	1	AE010357
29	37.6	36.6	143786	3	AC024859
30	37	36.6	298960	2	AC006896
31	36.8	36.4	104326	1	AB070940
32	36.6	36.2	27522	1	AB070942
33	36.6	36.2	107379	1	SHGCP1R
34	35.6	35.2	1044	6	AX195968
35	35.6	35.2	15970	1	AE007061
36	35.6	35.2	53450	1	MTV018
37	35.6	35.2	109519	6	AX195929
38	35.2	34.9	10670	1	SCARDIGN
39	35	34.7	12381	6	AX006889
40	35	34.7	24568	1	AF220951
41	35	34.7	30690	6	E38020
42	35	34.7	31422	6	E38021
43	35	34.7	47981	1	AF263245
44	35	34.7	47981	6	AX112026
45	35	34.7	50937	6	AR159871

ALIGNMENTS

RESULT 1
AF217189
LOCUS
DEFINITION
AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189.1 GI:7453554
SOURCE
POLYANGIUM CELLULOSUM.
ORGANISM
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epoIII gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649953
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epoIII gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submision
JOURNAL Submitted (16-DEC-1999) Kosen Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1..58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1..992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453353"
/translation="VRGHEAFADFGVPRVLLYDNLKNAVVERHGDAIRFHTLLAL
SADYFERPVPVAVARGNEKGRVERAIRVREGFEARAYADGLDRQATEWTSNAL
DSRWEDRARVQAQAFDDERSVLRHPDTPPDHEVEVEGCTPYARPDLDY SVH
DSTRRLVYADLSQVRIDAGNOIVATHYRSMRDGOIOEPHIORLVDEKRRARHR
GIDRLARARSQAFLRTIYAEGRDNGSAIARLLQILDVGAELDEAVYLEERTI
HGAVROYIDRRSERHPPVSIPTKGEHALVYTPHSLTYDLKQDP"
989..1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453356"
/translation="MTDLPTETTKDRLKSLGFLGLACMEQLADKFWLREVLAIEERE
RHRSLERLKRNSRYAAEFKPMTDGDSMPMKIDREAVDLDYSRYADLFEVYTRAYD
AOKPLILSTNRKAFADMGVFPAAACVYTLVDRLVHRAEVIETIEAESYRLKEAKELNAT
RTKORRTKKH"
1998..6263
/gene="epoA"
1998..6263
/gene="epoA"
/note="type I polyketide synthase; epoIII gene PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRIERAAEDPIAIVGACGRIPGVIDLSEFWTLLEGSRVY
GOVPERMDAAWFPDDDAPEKTFPTTRASFISDVAQFDASFFGISPRALMDPAHR
LLEVCWEALENAAIAPSAIVGTETGVFGIPSECEALPRATASAEIDHAGLDTM
PSVAGRISYVGLRGPCVAVDTAVSSSLVAHLACQSLRSGECSTALGVSGLMSP
STLWMLSTRATLATDRCRCAFAEADGFGFSGCAVAVYLKRLSGAARDRLIAVMS
SAIHDGASGGLTVPNGSVOETVLKRALADAGCAASVGVVAHGCTTGIPDETGA
LNAVYGLGRDVAATPLLIGSVKTNLSGPEYASITGLIKYVLSLQHGQIPAHIAAOLN
PRISMGDLRLVTRAKTPWPMNTPRAGVSSFGMSGTNAHVLEAPATCTPPAPE
RPAELLVLSARTAAALDAARLRDLHETVPSQCLGDAVFSLATRSAMERHLAAVAT
SSEGRALDAAGOCOTPGVYRGTIADSSRGRLATLFTGCAQTGLMGRLYDVMPAF
REAFDLVRLFNOELDRPLREYMAEPASVDALADOTAFQOPALFTFEYALALMRS
MGVEBELVAGHSIGELVACVAVGVSLEDAVFLVAARGMLALPAGGAMVSIAPAE
DYAAAVAPHAASVSTAAVNGPQVVIYAGGQPVYHIAAAMARGARTKLHSHAFHS
PLMAPLEAFGRVAESVSTRPSIVLVSNLSGAGTDEVSFGYWRHRAREVVRADG
VKALHAAGATFVEGPKSTLIGLVPACLPDAIRPALASSRAGRDPEATLEALGLM

AVGGLVSMAGLEPSSGGRVPLPTYPWQRERYWIDTKADDAANGDRRAPGAGDEVEKG
GAVRGDRSRARLDHPPESSGREKVEAAGDFPRLIEDPEGLDILVYRVERAPG
LGEVEIADVAGLSEFNDVOLALGMVDDLPKGNPNPLLIGGACAGIVAGVGNLVY
VGGPVALSAGAPATHTVTSALVYLPDPAOLCATNEAAMPVAYLTRAWYLDLAIARLP
GERVLIHAATGCVGLAAVOMAQHVGAEVHAITAGTPKRAVYLSLGRVYSDSRDFV
ADVRAWTGGEVDVNLNSLSELIDKFSHGHFEVLGRKDYCAQDLSRLPFLR
NLFSVLVDKGMMLERPARVRLFEILDLIAGVTPPPIATLPIARVADFRSAQ
AOLHGLVLTLDGPDEVOIRIPHAGAGPSTGDRDLIDRLASAPAPARAALAEFLAQ
VSOVLRTEPKVGAELPFRIGDLSIMAEVLNRRIPASIKLSTFTFSTPNIALLT
ONLIDLATATLSERVAENLRAGVQSDVSSGACQDDWETIAL"
6260..10492
/gene="epoB"
6260..10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epoIII gene PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINOLLNLEHGQVKIAADGRLDIOQAQKNAKNPLARISEH
KSTILMLNQRLPAESTIVPAPARHVPFLTDIOGSYWLGRGATFVPSGIAHREID
CTDLDVARLSRAFRKRVARHMDLRAHTLPDMQVLEPKVDADIEIIDLGLDRSTREA
RLVSLDAMSHRIYDTERPPLVHVAVRIDEEQOTRLVSLIDLINVDLSLITFKDML
SFYEDDETSLPVLEISYRDYVLAESRKSSEHQRSMQDWKRRVAELPPMLPMKAD
PSTLRIRFRHTEOMLPDSWSRLKORVGERGLTPPGVILAAFSVIGRWASPRPTL
NLTFRRLPVHPRVINDTGDTSMTLIDITDTRDSEFQRARTRIOELMEADHDYS
GIEVQREARVYGLIQGALFPVYLVSALQOVGYSIQRLCTPYTTSIQTLQDLIDH
OLYEHGDLVLAMDLYDGVFPDLDLMDLEAVALELRITTEEPWSBOMKSLPPQLE
ARASANETNSLSSEHTLGLFARVEQLPMQALVASARTLYEELSRSRRLGARLR
EGARANTLVAAYMEKMEQVAVVAIVLESQAAYPIDDLPAERLHYLLIDGVEKLY
LTOPMDGLTSMPPGIORLIDAGVEGGDLPMPPIOTPDPLVYITTSSTGLPK
GYMIDHRCQAVNTLIDINEREELIGDGRVALTSLSPDSVYQVFLIAGGTTIVPDA
SKURDPANHAALIEREKTVMNNSPALMMLVYBHSGRPDSLRSLRSLSGDMIPV
GLPEGLAIRPGVSYISLGGATASWTSIGTVRVNDPWSASIPGRPLRDTFVLVD
EALPEPPVWPGQLYIGVGLALGYRWDEKRNSEFLVHPETGERLYKGDLDGRYLPD
GNIEMGRDNOIKLGRVYELGEIEETLSKSNPNDAIVIVGPDANKLLAAVVP
EGTRRAAEODASLITERVDARHAAKADGLSDGEVOPFLARHGLRDLDKPAVVD
TGLVPEAGLIDVYARRRSVFLLEAPVPEVERGFLSCLSVPPGGAALPKRYSAG
SPYPOVTVAYAKSGREGVEGFEYHYHPEHEHLVSDGIERGHAHPNPNDEBA
RGLTVYGRIDATIESLYGSLSRPECLLEAGTMAQLMEQAPSCNIGVCYGOQDFEQR
PVLDLRHSDVYVHGMGLGRVDRQOVCTLGDODSSPRRATTCGAPGRQDRADLEID
FLRTKLPEYMPVTFVEYLEDALPLTSGKGRALRLRKDTSPRSRSHGAPRDALIEI
LVAAYVEVGLGVNGLOOSFVDLATSIIHVMSRLIQKRLDREIATITELFOYPLIGS
LASGLRDSKDLDEQRNMQDVRVARKKRRRS"
10639..16137
/gene="epoC"
10639..16137
/gene="epoC"
/note="type I polyketide synthase; epoIII gene PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEDESSAIVIGMSGRPPGARDLDEFRNLRDGEAVQRESE
ELAASGVDPALVDPYVAGSVLIEDVDFDAFCISIREALEMDPOHRTIMECAWE
ALENKGIDPTAEGSIGVYAGANMSYTLNSLNEHPAMKRWGMQOTLIGNKQDLAT
HVSYRLNLNGPSISVOTACSTSLVAVHLACMSLDRDECDMAAGITVARIPIRAQVY
AGCGITFSPGHRAPDAKANGCTIMNGCCVVLKPLDRLASGDGVRAYVILCSAGND
GARKIGETFPSEVGOAOLMEALALAGVARSIOYIEFGTSTLGDALIEYALRRV
DBDASRRSCAIGSVYTGHLGHEAAGIAGLTKTYLALHEHQLPSSLNESPSPSIDF
ASSPFTVNTSLDNWNTGSTPPRAAGVSEIGCTNAHVYLEEPAKRLPAAAPARABEL
FVVSAAALDAARLRDLQAHQIGLSLGVNAFLATRSRPMERLUPAAAPASSEAL
RGLDPAAGQPPGAVRGRCSPGNVKKVVFPOGSGVGMQVLALEEVFPAAL
SACDRAIOEAMNSLAEIADGSSQIEIRIDVYFALVAQALRAALRSGVGPDV
VIGHMSGEVAANHVAGALSIEDAVNLCIRSLRLRISOGGEMAYTELISIAEABAR
GYEDVSVAVNSRSTYLSGEPALIGVLSLANKVSGCRVKKVADVASHSPQDPLR
EDLLAAGGLRPAARAVPMRSTVTGAMVAGPELGANTYMNNTLROYRRAEVQADLOG
GHGLFVEMSPHPLITTSVEEMRAAQAGAAVGSILRQODEPRAULEALGLAMAGY
VPMGRILFPAAGRRVPLPYPMQREERYWIEAPKASAGRGRRVAGGHLGEMQTLST

QSTRMLETTLDLKRLPWLGDHRYOGAVVPPGAATLMAISSGAEALDGPLOTIDVV
LAELAFAGDAVLIQVYVTEOPSGRLFOFLOPSRGAHSFVHARGALLRVRTS
VPAGITLSAVBARLOASMPAAATYATLTEMLOIOPRAGIOALRMRGEGALGRILP
DAGSAATYRLRHPALLDACCFOYVGSIFAGGERTATWYEVGSLRLDPSSELNCHA
RVNHGRQTPRQCADENWVDSGVAVESVLVAQRLPGGVRREEDDWELELMEP
AAVGTAKYNAGRMKLLGGGGLGALRSMLEGGNAVHAESSTSAAGVRLAKAF
DGOAPTAVHLGSLDGGELDPGLGACGALAPRSADVSPALFDLALHVRQAOLIPA
OLAMGFRDAPRLMLTRGAOAVGAGVSTYQAPILGIRYIAMEHADLRCAVYDLD
PTPRDELGALLAELADDAEVALRGFERGVATVRROPTREGRILECVPTDYAT
IRADSTIYVTGSLGLGSVAGMLERAGHILVYGRSGASVEDRVAVALAEAGAR
VTVAKADVADRQOLERILREVTTSGLPKNVHAAGILLDDGLMOOTPARFEKVA
VOGALHILHATREAPLSEFVLYASGVGLSGOGGNAATFIDLALHVRQAOLIPA
LSVDMGLEAEVMAAODRGARLVSRGMSLTPDEGLSALARLEALHVRQAOLIPA
PRMWEIYPAAASSRMLSLVTAHRAASAGGPAQDGLRLRLAAAPRSASGLLEPLR
AOISQVRLRPGKILEVDAPLTSLGMSNLSMGLERIRIEMRITAPATLWYPTVAA
LSGHLARACCAAPESPHTTADSANVELEMSQDDITOLIAKFKALY"

gene
16134..37907
/gene="epod"
16134..37907
/note="type I polyketide synthase; epothilone PKs modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="WTRGPTAQQNPILKQAAIITDRLERLAGLAQAELEERTPIAV
GITCFPEGGADPEAFWELLDNERDAPDLDRMALVAVAPEAVPHMAGLITPEIDC
FDAEFFGISPREARSLDPQHRLLLEVAEGLEDAGIPRSIDSGTGVGAFADYA
RTVALPREERDAYSATGNMLSIAGRLSYTLGOLPCLTYDACSLSVLAHLCS
LRAGSDALAGVSTLSPDMMEAAATOLSPGRCRTPDASNGVGGGGLVY
LKRLSDAQRDDRIWALRGSAINHDSRTGILTANVAOETVAREALARSANVEAG
DYETHGCTSLGDPITYEALRATYGPARSQSTRVLANVKNIGHLAAGAVGLIK
AALSLTHERIPRNLMFRTLNPRILLESSALALATPEVWPTDRREFAGVSGFSGST
NAHVLEAPAVELEMPAPERSAELLVLSGSEGLAQOARLRHRLDHPFLGIDV
AFSLATRSAMTHRLAVTSREGLIALLSVAQOTPAAGARCIASSRGRLALFT
GOGAOTPMGRLGICAMPAREADPCRYTLDRLEDRLEPMEWMAEASAEILLDOT
AFOTPALFAVEYALTALRMSWGVPELLVHSGILVACVAGVSLDGVYVARG
RLMGSLSGAMVSLGAPAEVAAVAPHAANVSLAAYNGEQVIYAIEQAVQVIAA
GFPAAGVTRKRLVSHAFHSPLEPMLPEFGVAAVSVYRRPSVSLVNSLSKVITDE

Query Match
Best Local Similarity 100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 60
|||||
Db 33500 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 33559
|||||

QY 61 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCCCA 101
|||||
Db 33560 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCCCA 33600
|||||

RESULT 2
AR172664 AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE .
ORGANISM Unknown.
UNCLASSIFIED.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: us 6303342-A 2 16-OCT-2001;
FEATURES
location/Qualifiers
1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 101; DB 6; Length 71989;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 60
|||||
Db 33500 ATCGACGGTGTGCGGCGGAGCCGCGCTCGCAGAGTGTGCGCATCTTGCGGC 33559
|||||

QY 61 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCCCA 101
|||||
Db 33560 AAAGGGGGTGTCTGCGCGTCAAGGTGAGCGTCCCA 33600
|||||

RESULT 3
AF210843 AF210843 68750 bp DNA linear BCT 21-JAN-2000
LOCUS Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
DEFINITION cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE .
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar,I., Schnupp,T., Ono,M., Zirkle,R., Milamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Strutzmann,A., Cyr,D.D.,
Goriach,J., Mayo,J.M., Hu,A., Goff,S., Schmidt,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microbubble-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
location/Qualifiers
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDLFAFARTSARVLDGHRATERHVLAERGIEDLRALREH
LRIOGSPSHCMICGLLVELLADHODPLASISPHARSIRLHPDTSAMLVDPALV
RWLARAGAPRLREYERERARKTOEARRLMLAAPCEPADLPRFDDNANGELG
MSFEVAERLRASITAPBELACALMLMLGAGPWSGITPAYELPENLIGELPLT
ALAAASAPGTSERLARGARLFAVSEVVSRSKSGQNGNIPDALMERLRTIVRAGNAD
LSRPERARLAEVRLRROPAPFAAGGLAVAGVSSGRSLGVLTDALVSGDND
LYMFOGRISPVLLAGNDPPELAPLSOMFLVHANAGTISKYLTGSPILVARN
QARMSIVHARGFMAVNOAMPDEBERGAPFYVORSTIMEFHPTRCLHEPAGSFA
LACDEHILWCSLGRLEIMHPHRRGAPSRAPVAGENRPAATWYPSLTNATHVL
WADPRRAILGVDRKTEVPTVLAETRRPRAVSESDRIRALTQDPDSRDMVHEHR
SGASTVADYQQLMDRPDMVLRNRGLTFETINDRLITLARS"

CDS
1..71989
/organism="unknown"
/product="unknown"
/protein_id="AAF26905.1"

CDS
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"

Best Local Similarity 100.0%; Pred. No. 6,5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 61
|||||
Db 39113 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 101
|||||
Db 39173 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 39212
|||||

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6,5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 61
|||||
Db 39113 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 101
|||||
Db 39173 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 39212
|||||

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6,5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 61
|||||
Db 39113 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 101
|||||
Db 39173 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 39212
|||||

Db 39113 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 39172
OY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 101
|||||
Db 39173 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 39212
|||||

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6,5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 61
|||||
Db 39113 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 101
|||||
Db 39173 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 39212
|||||

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6,5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 61
|||||
Db 39113 TCAGAGGTGCTGCGGGGAGCCGCGACGCTCGAGAGGTGCTGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 101
|||||
Db 39173 AAGGGGTGTTCTGCGCTCGAGTCAAGGTGAGACGTGCGCA 39212
|||||

```

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.
JOURNAL Genes for the biosynthesis of epoethilones
FEATURES
source
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCACAGGTGCTGGCGGCGAGCCGCGCTGCGCAGAGGTGCTGGCGATCCTTGGCGCA 61
|||||
Db 39113 TCACAGGTGCTGGCGGCGAGCCGCGCTGCGCAGAGGTGCTGGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTTGTTGCGCGTGCAGTCAAGGTGAGCTGCGCA 101
|||||
Db 39173 AAGGGGTTGTTGCGCGTGCAGTCAAGGTGAGCTGCGCA 39212
|||||

RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.
JOURNAL Genes for the biosynthesis of epoethilones
FEATURES
source
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 99.0%; Score 100; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 6.5e-14;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TCACAGGTGCTGGCGGCGAGCCGCGCTGCGCAGAGGTGCTGGCGATCCTTGGCGCA 61
|||||
Db 39113 TCACAGGTGCTGGCGGCGAGCCGCGCTGCGCAGAGGTGCTGGCGATCCTTGGCGCA 39172
|||||

OY 62 AAGGGGTTGTTGCGCGTGCAGTCAAGGTGAGCTGCGCA 101
|||||
Db 39173 AAGGGGTTGTTGCGCGTGCAGTCAAGGTGAGCTGCGCA 39212
|||||

RESULT 10
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000

```

```

DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 10910)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
BIOTECHNOLOG FORSCHUNG GMBH (DB)
FEATURES
source
BASE COUNT 1852 a 3124 c 3613 g 2321 t
ORIGIN
Query Match 72.1%; Score 72.8; DB 1; Length 10910;
Best Local Similarity 83.0%; Pred. No. 1.7e-07;
Matches 83; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 2 TCACAGGTGCTGGCGGCGAGCCGCGCTGCGCAGAGGTGCTGGCGATCCTTGGCGCA 61
|||||
Db 4323 TCACAGGTGCTGGCGGCGAGCCGCGCTGCGCAGAGGTGCTGGCGATCCTTGGCGCA 4382
|||||

OY 62 AAGGGGTTGTTGCGCGTGCAGTCAAGGTGAGCTGCGCA 101
|||||
Db 4383 AAGGGGTTGTTGCGCGTGCAGTCAAGGTGAGCTGCGCA 4422
|||||

RESULT 12
LOCUS SAU421825 66808 bp DNA linear BCT 16-MAR-2002
DEFINITION
Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
ORF1, stIIa gene, stIIb gene, stIIc gene, stIId gene, stIIe gene, stIIg
gene, stIIg gene, stIIh gene, stIIi gene, stIIj gene, stIIk gene and

```


ORF9.
A421825
A421825.1 GI:19572309
acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
ORF7; ORF8; ORF9; polynucleotide adenyltransferase; potassium
channel beta chain; ribosome binding factor; stia gene; stia
protein; stib gene; stib protein; stic gene; stic protein; stid
gene; stid protein; stie gene; stie protein; stif gene; stif
protein; stig gene; stig protein; stih gene; stih protein; stii
gene; stii protein; stik gene; stik protein; stil gene.
Stigmatella aurantiaca
Stigmatella aurantiaca
Myxococcales; Cystobacterineae; delta subdivision; Myxobacteria;
Bacteria; Proteobacteria; Cystobacteraceae; Stigmatella.
REFERENCE
AUTHORS
TITLE
Gaitatzis,N., Silakowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
Hofle,G. and Muller,R.
The biosynthesis of the aromatic myxobacterial electron transport
inhibitor stigmatellin is directed by a novel type of modular
polyketide synthase
Online Publication
JOURNAL
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
Braunschweig, GERMANY
FEATURES
source
Location/Qualifiers
1. .66808
/organism="Stigmatella aurantiaca"
/strain="Sg a15"
/db_xref="taxon:41"
complement(4978. .6231)
/note="ORF8"
/codon_start=1
/transl_table=1
/product="cytochrome P450 monooxygenase"
/protein_id="CAD19077.1"
/db_xref="GI:19572310"
/transl_table=1
/product="protein kinase"
/protein_id="CAD19078.1"
/db_xref="GI:19572311"
/translation="MEELISALPDGSRVGPWRIEGRAGYGYGAVYRAHVGRICDR
SVLTKARVNDLAREBEARELLEVRANPRLIGHSTWRAGPEGEAHPYVYMWYEG
VRLQMKRERLRTAREVLAQVAPALAEATFHMHKGLRDVKGENVLVSPEGASLMD
FGCGTWGAAPLTGGLAPGTAKAYSPBALFPHNHNNEADAHYQATPADVDALVGT
AYRLCTGIVPPATPEGVGERETQEVKKPPSQVPLPALPOLILRMJRDPKER
GRAGELAAEMALERAAGADEPMNPNHSVPSEITGRPVSVOREMGMPVALVLA
VGLLMSGSLNTRRYAPASBMTGGTGADAAVALPVSAAAREPSPGGLMDL
KEPLPGRRPPRCFQISIOGCGWLEVPQSSPGCESTYEMKGCYSPVAAPPRIS
DPO"
complement(7849. .8577)
/note="ORF6"
/codon_start=1
/transl_table=1
/product="methyl transferase"
/protein_id="CAD19079.1"
/db_xref="GI:19572312"
/translation="MSONIVDDPFFESYSOLASLKLGLAGAPMPALRALPDVKDL
RVYDLGCGFWCFMARKNDAAHVLAHGLSEKMLTRAKAETSGDAITYQOADLERVL

PGAFDLAFSSLAHYVEELEPLLATVHRGLTPGWFVFSIENPFIENASQPEWMTSP
EGHRTPEVDSYOLEGKRTTHWLAEEVVKYHRTLGTTNALLEGFTLRHIEGMPDTDA
QLSAHPLEAERDRPMFLVSAOR"
complement(8622. .8879)
/note="OR5"
/codon_start=1
/transl_table=1
/product="acyl-CoA binding protein"
/protein_id="CAD19080.1"
/db_xref="GI:19572313"
/translation="MALEDPRSQAERKALKTRPSDRDLLEYLSLEKQATEGDVQK
RPGILDYKERAKYDAMWGRKGLGRESAMQVVALVERLRG"
8961. .10070
/note="ORF4"
/codon_start=1
/transl_table=1
/product="potassium channel beta chain"
/protein_id="CAD19081.1"
/db_xref="GI:19572314"
/translation="MYYLICRSQSOLRWILNGSAHVRFPANGARPRMRHFRHGRS
GLMYSISYGNMLTHGSOVEERAAVACVRALDGGITTPDADYACTRAEYVGRAL
OGORBELEITFTKYWPTGPPNDRCISKRHITESHGLSRKLTETVEDLQAHRYT
ETPLEETMQAFSDVVRQGAHYIGVSEMTADQREGVKLSKOMGFOLVSSPOYSMLM
RVIAEAVVPASQELGVGVYSPYAGQVLTGKRPQKPEGSRATDEKGSASFIRF
LNDVLTFRVQDLTPVAOEGAGLSLAQLAVAWLQPNVSSAIIIGASRPBQYKENVKAAG
VKLSEEMRHIDTLGPIVERNPCKTLENSPRKH"
10428. .10862
/note="OR3"
/codon_start=1
/transl_table=1
/product="ribosome binding factor"
/protein_id="CAD19082.1"
/db_xref="GI:19572315"
/translation="WTSFKORRRAPRGRTSSLSSEDAOSLPAAHILRVOSTLFOEVA
RLFRGLSDPLLEGVSVTSEFLTPDGNARLGFLEPYEAGSAPLEALVYRASGPI
RAQLALGILNLRVPHLRFVCGVAPPGFSIFEDGSAEEGGS"
10859. .12016
/note="ORF2"
/codon_start=1
/transl_table=1
/protein_id="CAD19083.1"
/db_xref="GI:19572316"
/translation="MRPRRIGAPAVPILAMARAVAGAVKOLLRAPYLAEYVA
AMPVHAAGVAGVATVPRATERHIVPGALGGGCGSVAVRPDPFMAISLGRVLRLLA
AFARAVYGDVHNGRGSLSPGGLDIGHLSHRLQREMERLYPKHGLTSGGNHFL
DRDEGANRLDILHTRGSGVGAIAAHHLYAQSVEGSLPELRTDTSEGAACLADTEL
ACQFARANRLDILORAVEYLAGAIGQSPVLDSDRVVHNNHVAASHRGRLVARKGA
VGLVAGQGLIPGSMGTASYLVAGKGFQAFSCSHGAGVMTFRERARERIRPALAH
ALRRVFDGGRVTLVEAPAYRDITEVLEDEALVTPLRRLPIAVLNG"
complement(112082. .13974)
/note="ORF1"
/codon_start=1
/transl_table=1
/product="cellulase"
/protein_id="CAD19084.1"
/db_xref="GI:19572317"
/translation="MKSLLMLMLAVPLVGTALARPANAREPGFEEPPENAGLTA
AQCNPVPLSTRGYIVDRCGERKLKSVNMFSGSDOLEVVGGLDKLLEDAVAGMA
MGFNSVPLPSNNKLRHPDNDKPAQORLEKLSVCTDKRPELNTLNTALVDAVVA
ELTRQGLVILINHTTSSMCCGMDGNGFMSQALROMODMWMLAARYKQKMGV
ADLENERPDLDSPPNMGHNMHMAQOMGLLRLTNPDDLIVEAIVANMGLDLS
SRPOLKPVORPVALRGDKLYVYVHNYGTGAPNQSGLSGSGKYSDDMKPTLYGLT
DOENGFALANQATTAAYWMSSEFVGVNBOAANSRANFSLADYLIEKDDWATWAIN
AAKLQNSVVGEDETTYGLMSTYPMNGVNSGDWRLTGRLGADGTGAVDATRFLP
MTVLLKDGSSSTYVDNNSLDSDSHSGRAKLSCPRGYMWGVSAFTLLCINAGAVPAQ
OGTGSYOTITQOEVSLPRYPGDMANOSIKFEDPGTGYVAGISTANPFLLELAGLICEQ
TGGLPLNSKSAKNFWGDQRRASSSGGDMVSQYKGCQADNOYLIGLAHRRSGADYPS
VVICSN"
14345. .21466
/gene="stia"
/db_xref="stia"
/function="PKS"

```
/codon_start=1
/transl_table=11
/product="Stia protein"
/protein_id="CAD19085.1"
/db_xref="GI:19572318"
/translation="MTQLASRLGVGSESIDPERSRRTGIDSDRTSLLAELGALLK
PLSPYLMTVPTVEALSRHLAGADVSGGSDADTLDVDDPDAIIIGMAGRPPKADY
ESFWSLLRDGVATITEVPPERMDLAFYGVGLSPKMTNMGFLDQVDDPDPFG
ISPREAIQMDPOOLRLLELSEWALEESAGVADSLMSRGVSGAMSSYAKLTAL
GLGPHATGODSSIIAARVSFELGRPSLVNTGSCSALVAHLACGSLRNGEAM
ALAGVNLISPESSVAMSOLGALSPDGKAFDARANGVGEVGLTDPISAT
ANGPYLCIRSSAVNNDPFSNGLTPNRADEYVLRDVAARAGSPDVHYETGT
GTVDGPIEALNLSAVGVNPAERPLIIISVKTNGHLEASGLNGLKVVLSIQR
ALPPNLFEPQPNPYIRFELRLVLPKHPWNEGSALADGISEFGGTNCHLVIA
PSGSHIAPLAETPEELSLARQARDVSVVSGSEVLAIDCFVAAACRESQCLA
LIARPAELAAOVDFLAGEKRGRLRGVTSQSKPVEVSGSGVLAIOYGLAM
TLPFRGATREGRDRLRGISGMSLEELSGAKSPSCSGLOTPIFYONGIATM
RSMGTEPACVGVSHICEVAHAHAGTILNEDARVYHHSRLMGRISGGAMIVDLS
AEEANRLREWEHVVVAACNAPASTVLSGAREVIELAELEERRGIAHYARIDTPS
HSPMLDFMGELEAVLAGLSPRVAAPNVTYGAELKAEECTAAWVRNLREPVLS
QALTKLEQDLRLLELSPHILARSVEQCLGEGRGVLAISMRGENEGACMRESL
GRLEFGTGVAMQALYPAKAKRYELPMGIGTSLDELGSRALRLPSHTAEALDA
ARALASITAKHPDRLDLCYSASVGSVHAHRLAPVTDKRALVHLDDPLKESRP
GRAGCATGEPKRYVFPFGQSQOMGQROLREPVFREAIETECDAIITSACMS
LIDADADEPHSRISIEVITPLFAIEVALALMSWGVDPDAVGHSMGETAACV
AGALRLTDAARVYCERSQLLRVAGGGALVELSLEQAREMISRDGRVSGIASNP
RSTVLSGDAIALEILRELEAQTGTCRKYVAVASPOVDLKDILLVETIAPRA
ATVPLSTVTGTEIGARMDASMMNRNLREPVLEGAIERLGGDGVGVLELSPHIL
LPSIDQTLRKHSHKGVLPRLREPERAVLIESIGTLFTGGGRVDFRHLPLGHRV
RLPHLPMORRWMIDVANAACASVGAAMHGGGSEHPLHLRLPSLAHAARCHWE
NKLRSASVSLDDHRYGVTVLPSTISMAAAGSGKKTVSILADLEKMSFL
ETASRTVQSLTIOEAGHSSFLRHGHSSTDEWTLHGGETVHVEEDGPPSADPVLA
AVEYARLPALDELGLQALIVLAHEGCEGLAIPTRIESIRAHSPGARMCVRLPVE
LEBEYRHPALDELGLQALIVLAHEGCEGLAIPTRIESIRAHSPGARMCVRLPVE
RRAASRGEVLEFDESQRLELOGARQIILDESGRVRVQMLHEVMPAPV
VPTAESSGGRVLEFADGTVAGELAAQIARAGEPCTIRYIPKMAPREADHRYNDE
OPEVARKLFGALSRSTSTIPRLGIVHMSLDAPSPEGVDALSRARLDSGSVLSLQ
GLNAELAQPRMLVLTGVOAVGASGAPGAVNOPLNGLOVAVAEALRYCTVD
LAPVSPGEEQTLQELLPADDDOVALGRHVRHLRCTRDEVAGSSTDLRACAI
RSDGCLITLGGAGLGLITIAKMLVARGAHLVLVGRGASDAEILAGMRGASVACL
VMKADVANTSEVEBELVGTIOKMPPLRGIFHAAGLDDGTVLQTOARLSVAPVPE
GAMNLRATGLASIDPEFVLFSSAASVSGSLPCNGVANAAGFIDALAGLRKAGLCS
IOMGRSEIGLAEODNRCERLALRGMISLNPREGIEALEALISPEEPREARITVL
RIDARWCKSTPAIRSSPLFSVLAGEANAKOYLERGSLRANLEPAPTRRALESF
LRBEAAQVLRPPARIDLSRSLQTLGMDLSLLELRNREANLELTPVIMVNPPL
SALAEPLAERMGIQLVGVEPPOAVESPSVSEBDGAMLMKEQFOLSNEDLGLMS
DDEDEVSS"
gene
21463..26220
/feature="stib"
CDS
21463..26220
```

Query Match 57.4%; Score 58; DB 1; Length 66808;
Best Local Similarity 74.5%; Pred. No. 0.0004;
Matches 73; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
QY 2 TCGACGCTGCTGGCGGCGACGCGGCGAGCTGCTGGCGATCTTGGCGCA 61
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 39912 TCGACGCTGCTGGCGGCGATCTTCAAGCGCCCTGCGCACGCTGAGATCG 39971
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 62 AAGGGGTGTTCTGCGTCGATCAAGGTGAGCTGCC 99
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 39972 CGAGGGGTGTTCTGCGCGCGGATCAAGGTGAGCTGCC 40009
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

RESULT 13
LOCUS 188042 49377 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 1 from patent US 5716849.
ACCESSION 188042
VERSION 188042.1 GI:3407982
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

```
REFERENCE 1 (bases 1 to 49377)
AUTHORS Ligon,J.M., Schnupp,T., Beck,J.J., Hill,D.S., Neff,S. and Ryals,J.A.
TITLE Genes for the biosynthesis of soraphen
JOURNAL Patent: US 5716849-A 1 10-FEB-1998;
FEATURES
source
1..49377
/organism="unknown"
BASE COUNT 7247 a 19522 c 14477 g 8131 t
ORIGIN
Query Match 52.1%; Score 52.6; DB 6; Length 49377;
Best Local Similarity 70.7%; Pred. No. 0.0074;
Matches 70; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 2 TCGACGCTGCTGGCGGCGACGCGGCGAGCTGCTGGCGATCTTGGCGCA 61
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 3300 TCAACGCTGCTGGCGGCGGCGGCGAGCGCTGACGCGCTGCTCCAGGCACTGGAGCG 3359
||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
QY 62 AAGGGGTGTTCTGCGTCGATCAAGGTGAGCTGCC 100
||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 3360 AAGAGCTGTTTGTGTCGATGCGGATGAGCTGCC 3398
||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
RESULT 14
LOCUS SCU24241 67523 bp DNA linear BCT 24-APR-2002
DEFINITION Sorangium cellulosum acyl-CoA dehydrogenase (sore),
beta-hydroxybutyryl-CoA dehydrogenase (sore),
methoxymalonyl-CoA synthase (sorc), reductase (sora), soraphen
polyketide synthase A (sora), soraphen polyketide synthase B
(sorb), O-methyltransferase (sorm), beta-mannanase, and
xylanase-arabinofuranosidase bifunctional enzyme genes, complete
cds; and unknown genes.
ACCESSION U24241.2 GI:13346872
VERSION
KEYWORDS
SOURCE
ORGANISM Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Mycococcales; Soranginiaceae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 67523)
Beck,J.J.
Direct Submission
Submitted (07-APR-1995) Ciba-Geigy Corporation, 3054 Cornwalls
Road, Research Triangle Park, NC 27709, USA
REFERENCE 2 (bases 52293 to 58665)
Schnupp,T., Toupet,C., Cluzel,B., Neff,S., Hill,S., Beck,J.J. and
Ligon,J.M.
A Sorangium cellulosum (myxobacterium) gene cluster for the
biosynthesis of the macrolide antibiotic soraphen A: cloning,
characterization, and homology to polyketide synthase genes from
actinomycetes
JOURNAL J. Bacteriol. 177 (13), 3673-3679 (1995)
MEDLINE 95325306
PUBMED 7601830
REFERENCE 3 (bases 1 to 67523)
Ligon,J., Hill,S., Beck,J.J., Zirkle,R., Molnar,I., Zawodny,J.,
Money,S. and Schnupp,T.
Characterization of the biosynthetic gene cluster for the
antifungal polyketide soraphen A from Sorangium cellulosum So ce26
Gene 285 (1-2), 257-267 (2002)
JOURNAL 22035375
MEDLINE 12039053
PUBMED
REFERENCE 4 (bases 1 to 67523)
Ligon,J.M., Hill,S., Beck,J., Zirkle,R., Molnar,I., Zawodny,J.,
Money,S. and Schnupp,T.
Direct Submission
Submitted (27-NOV-2000) Natural Product Genetics, Syngenta, 3054
Cornwalls Road, Research Triangle Park, NC 27709, USA
REMARK Sequence update by submitter
COMMENT On Mar 15, 2001 this sequence version replaced gi:1868150.
FEATURES
Location/Qualifiers
```

source	1. .67523 /organism="Polyanquium cellulosum" /strain="So ce26" /db_xref="taxon:56" 57. .986 /note="hypothetical protein; Orf1" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="unknown" /protein_id="AAK19886.1" /db_xref="GI:13346877" /translation="MRADNPSSAGICDAIERIYEHQGVLAAGRTTFLHYLEHGVSPA LFRPAEAFAVEEVEGREGRAAFASIGEKSRSPSPAQAPASSLSAOCRIAPADEPL NELCEIYRDCVATSOGCEERLAEMYGOPVPLSVQEMHEANLEAOCRIAPADEPL ADHLRAMGVDPSSMSAPAYPVNTGSAOCTEPPLPPRQQAADRPAPHSSEPT HTGPALKLLAVAKREFGVMLGTYGAHGLRRVRQDEIKRTNMPEDVQVAVLQHWISAG KSAPGPPLSDGELSAVPBARITRSPKMMRR" 1424. .6127 /note="hypothetical protein; Orf2" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="unknown" /protein_id="AAK19887.1" /db_xref="GI:13346878" /translation="MRSFPRKFNPSDPTWAGPRAATPRHREDAFOQLVGNALREFFA GVRYMPTEGDGIIVLENDARLPFVSRLDAGVIVECKDNDARGRATENVRAAK KVEEKLQVSAAGMPTGYRPMWATRYLYLTLSALPQDARNRLQHDVGFQRLRD GRTEIEHPVADWTDLCSLDRNPRILADRWLGLRSLRSQAQLASLGGVSRYLVD PTAPGRESQTHPDRLLAHYLERAGDEGRVLLVAGAGVGTRTLLEVGQADAGMRY LHVLSGEPPLSSDDLQOEYVYGRDRTLVICDYDPORALDVYSIRORLLPAAROGFV AIIASAATPDARSRNPRPEFAYRRLALETEQOARLACTMEQOIAPOAYMGLGEKX KLTDQPMIALLVKLELERLAATRKILTRAODITVREGLNHLERRLAHDELADGAS SRWDEKTSVILGAALALAFCPARKHVAVITCVLAENKAAADQAEIYVNLIEL GRNMAEGLTVLADHVVTDQLLEOALFTTSDAVRNALRLRLPCASPVILAFATA LRRVIAAGGFPMRELERQAQWLKASPALGRMLAGASAKSSAALMAMRGSPWKOT TTACMDQIVAPMLVHRGASSDAMPPLALCIRKNLREPPVAAALAEALIMVTEOHDEE HAASITMLIDRPDGSALISRLPELAMPYERHROPSAALVIGLLERPDLMARLAI ERLGPETLRWITEPREYEAGLPLGMLRREDTSSALSQVEGASWEMLERHROTA GYVHLGLLTCDEESASAKRAVDALMLKLIHADVAGGLVLSLLIRGSIPEQPLA VIDAAMWFKRRRRSPFVSLPGLLTTSALDDPRFQAQKIALGVAAADRYPOARI LVHLVSRGDLDEPAGALDREACSILQKQPDMSLSLGLMLEGTLPPDDYARVA ELALGWLERNVGEGLHVLVATSIACPHVDVTPARLAALAOAESGAVRNDGAVL GSIIRHIGITSLORASAOQMPGSHADLAIOAYRLLLRPDLDRLMARLYG AOJLMEOHGMCAVAVIAPILAREDTPEITRAEPTTATTLLEQHNPREAVDILNV LATSRITNKPGISIVAAAREWLQMSRRPLASSLIANLIPSAEPPSPDIDHALSW LEAHGDKPDAFDILLRLSARDLSPPRASVYVSTGLMSAHPHAELIADLIYTLAYR PEIDASVYVOLVSRAATMDLTHLSEASGALAAVMPAAORPELAHSVGRPLGAV EMLRDHASTHAGPVATILYSRKNITPTTLAATATSVKOMLGHAHREESTLPLAAL OHRNIPGLARRYVOKAVEIDHHTHYGDAAVAILHELFGDLIDDEPRITYFLAALRL GEHYDTPEGQQLRLVFSQASPESSGATECODVESLVDTLASSVEGEGEDCRALP ALLVEARSRAPALERRARVLVARALDRMRHRLRRRMAGACORLLEADAMPDRARA LQLLSEIGLPALED"
gene	complement(7505. .8896) /gene="sorc" complement(7505. .8896) /gene="sorc" /codon_start=1 /transl_table=11 /product="acyl-CoA dehydrogenase" /protein_id="AAK19892.1" /db_xref="GI:13346883"
CDS	/translation="MPLSLAPHEHARVERQGVDEVAVPADAFHRAQRTPEELIAL LAERGYLGISIPKEFGAGEDAVTLGLLAELAGCSVSRLTLVHTVMALAILRWET RAORERMLPELAGRRIGALALSEPVGSDAKSVTRTIAAGTDLVITDGKRWITGEE IADLFTVGRTEEGPAVLVERGPRGLTEPIDLIGIRASMTASVHLDCCREADHY VGRRLGIGITQVAVSALDILGRYTVAMGCVGILRAQVEACVATYSREOFGPIKHEH RMLISMTDLHASEMCMACAGLRODRDGCALSATMAKYFASWTIDKSGSEBPYAT AYMPTRLGFPARGAAKKAPLGTAAASFHASPLPRDRSRSTATSPRNGSLAS TVIGAPORRPAARLUPROTSTSPSELIRPRDRDPALLEVKKPGGYKWMRDHGPAG VVAVGSSGSFSK"
gene	complement(9199. .10080)
CDS	/gene="sord" complement(9199. .10080) /gene="sord" /codon_start=1 /transl_table=11 /product="beta-hydroxybutyryl-CoA dehydrogenase" /protein_id="AAK19885.1" /db_xref="GI:13346876" /translation="MDRSALFQPVYGVGAVGVGVAOSTAQTGHDVYLIVDSEAA LARAKGIRNGLAVALTLFSADEKRRAGDKRAVLIERAFITTDIGRLAGAPVBNYTE KWDIKREVVARLEGVCRPEIIFADTSAISITRIGSVTKRPSOVGMHFNKPVLEKM VSEVTRGFTHSPETEGAARFLAEKGKCVVVEAPGASVNRVLMLTINELIFLVQDV ASADVDIRIFKGCENHGMPLFETADLIGDLTIYSILEVYESFMDKXRYRCPILKKMY DAGLHGRKSGRGFTYE"
gene	complement(10080. .12671) /gene="sorc" complement(10080. .12671) /gene="sorc" /note="Sorc" /codon_start=1 /transl_table=11 /product="putative methoxymalonyl-CoA synthase" /protein_id="AAK19884.1" /db_xref="GI:13346875" /translation="MTPTALHETNGSPQNSLOISEPLRLALVAEGSSAEFLTASMDGLT RLTAAAGSDVIVIRVOPPGARAPASHRGTLLDDATKTVARRETPAAGAMPPEAVFLP GLGDHYGMARDLITNALPTFREQYDRCAALLITPELGVDITRYMHAKRASAARCAEL DLRMLERKEBGSPEERRLNQARHAPALVEYALAEQWRAMGVLPDMLGISLGE VPAACVAGVSELEDSIALVARRELIESLPAGMALVMLEEETLPQGLHLSVAVN GPBECVAGVLEVEQVLEARLIDPRVYLRRVQSSHAFSSHMEETIERVRIYAGFTY RPPTPEVSNVTGTLITPAEATDPAYVAIHLCRPVADGLRALGSVLELVPGQTL LSLATLARKAGAPSSGSSPTPARTIYASMRHAYESRDVSVILKAYGRMLSGAATE GEOPATISAEAPARADAPAGCSNAPVYEIISPAEPAPLPTDTEREVAALMOTLLRE PPTDANPFRUGGSLATRLIDVSSLSLVKQLPLRVYOTSTLADAAKIDIGLRHG ARBPAAVOGSAAPVNTSAPTPAARPAAPADAPAPLPLRPLNGLSVAHQAEFRHY DDIFAHNSYVKGHGRIDGACVLVDGNGILFTLFAETEKADVFETFEAPPLFELL SRNALGVATLLENGLSDRENDAPPTFYPRSGMSGFYDDEAEENHLKAIMDNOR RGGAAYVAQMDGTATEELLDYREAPVYFATLRLRSLDREORIERVDMIDVQKCL YLEGIDDAWMPKFSQVLVAHDAHDSVALLSLFENHGKVIYAEQDELVGNINIRKI YALREGV"
misc_feature	complement(11316. .12374) /gene="sorc" /note="Region: A1 domain" complement(11316. .12374) /gene="sorc" /note="Region: ACP domain" complement(10083. .10880) /gene="sorc" /note="Region: M1 domain" 13697. .14158 /gene="sorc" 13697. .14158 /gene="sorc" /codon_start=1 /transl_table=11 /product="reductase" /protein_id="AAK19893.1" /db_xref="GI:13346884"
gene	/translation="MSVLIHFFTRFRPPIRIDAPDTPSSERASMLAQYRSGCTPMYGV DAFLFSSSAGVLGSGLOGAVASANAFLDIVSEPRVIGRSATSIAMCINNGGIDYDH AGKHLRRGVSVMPLDLALAAVSREBSRGSSPQRSSPISCSIPVSSRCBG"
CDS	14325. .33272 /gene="sora" 14325. .33272 /gene="sora" /codon_start=1 /transl_table=11 /product="soraphen polyketide synthase A" /protein_id="AAK19883.1" /db_xref="GI:13346874"
gene	/translation="MTKEVYRPGSAPLTGEGDLTLIVAHIAERLMDARFIDVHEPFS RHGDSGAVDLVDLRTALGRPLSPVYVVOHPDPAIARHLAGGADARBGQARADSA YERGAENPELIAIGMACRPPGAPVDVSYWRLLSGVDATVTEVPAGRWMDAFYDRBP

gene
CDS
IARLSEIAGGAPLRGVIIHAAGVLDGVLLOQTVERFRVMAKVLGVNLIHVLTHE
APLDFEVLYSASAMFGAGPGAGVAGNAFIDALAHRRALGDLGLSINNGPEFVGL
AAAGANGERLARGSDSLTPAGEENTILGRLLDQDTOMAMVPLELRKVEYPRAS
SPMLSELVPADAGCEVGEARDVALLETLRTRTPDEAKAALEQVYREGOLARLRDSR
IDEPALQSGFLDSLMGLELRNLVSGVSLPSLTIWKHTPLAALCAHLGEMDRP
LAETLAGAAGAPVADESAADKEVFL"

complement (11487..14303)
/gene="maxb2"
complement (11487..14303)
/gene="maxb2"
/note="PKS; dehydratase: enoylreductase; ketoreductase;
ACP"

gene
CDS
/codon_start=1
/transl_table=11
/product="Mxab2"
/protein_id="AAK57186.1"
/db_xref="GI:14210839"
/translation="MSKOEKPAEQPSALQRAALIVEKQARLDAVERATEPIAVIGL
GCRFGGAVDSSSYWKILRDGVDALREVPSSRMVDPGHFSTRGVPGKMTGRGFLD
DVEHDFEFGISPREAASLDPOORLYLEAMELENAGLADPLDAGSKTGVGWS
SDYMARLKENDAIRPDGMATNGYSFVGRVSYVYLGQPCMPVPTACSSSLVAH
LASESLRAGSNTALAGCVNLISPEMTICLSKQALASDRCKTDAADGVYREGS
CGVLVLRSLDAQRDGDSILALIRGSAVNHDGASGLTPVNGPQQAQVQALDNARI
APALVGYIEAHGTGTPLGDPTELRLAGVAVGKRPADRPFGISVKTNIHLEPAAGI
AGVYKTLISQHKIEIPHLHERPNNPEMDRIIPARVPERVPMWAEGRRIAGVSF
GLSGINAHVYLEEPASTAPVAEEDKQALVYLSAKTPRALSLEAMSSFLKNEEG
GSLADISYTSALSTPREHRLALAASKOEILRLEAPAGKSHAGLSGSRGTGLCO
KIVFVFGGGSOMLGMGQRLIAENVTFREAIIDWCHEAISRITDLSRLDILTNPGQPR
WMDIDVLTPLFALQVLAALVAVMWLGLEPAVGHSMGEVAAAHVAGALSLDPAALII
CERSKILRRVSGGAMAVVDLSREQAEVIRGAEDRIAIVSNGRSTVLSGDPAALII
EVLRLERREVFCRWKVDVAHSPOMDPLRODLDBMAAIIPTGSAVPLYSTVGAP
IDGRLESAFVWVRNLRQPLFADAVORLIDQGVNLTIELSPHPLTVPEVDAMLREGDT
SGRGLVPSLIRDOEEMPSLISVGALEFVYIANTGSGYIPTVSAGLRPLTRGSASG
TGNSCOOLRRRAGSCALARKEVTS"
complement (14300..29878)
/gene="maxc"
complement (14300..29878)
/gene="maxc"
/note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
ACP regions"
/codon_start=1
/transl_table=11
/product="Mxac"
/protein_id="AAK57187.1"
/db_xref="GI:14210840"

ACRRPOADNVEAFWRLLQGVDAVREIPLERWPKDEVADIGDAMRWGFLDSVSD
PGFEGISPREAVRMDPOORLLLEVAMEALDAGLDVXLGSRSGVIGACNDYHGM
OVERPETGDAFSATGVAASVLSGRSLYLFNLGGPSLVVDACSSLSVLIACOSLRN
RCNMALAGVNLILSPSIVLSKIQAISPDRSKAFDASANGFTRGCGCYTLER
LSDALADGDHIIATIRGSAINODCKSTGLTTPNYLSQALIRQALLESSGLAKBOVSTV
EAHGTTGLDPTLEALRETYGVPRPDGSCVCGISVKTWGHLSAAGIAGIMKVV
AMRHQSIPLHLHLKQVNPRIQLBGSALTPTRLTPMKASGOPRRAGVSFSGISCTNAH
VILEEAPAPYAKDALPQRPVPLPSARSRLVALAERYAERLREGGSLDVCYTA
SARHTHTHRLAVLADSRHTAERLRAFAREGDPSGAMSGKAGRRKIYFVYPOG
AORAGMRALMESPPVRDALIVYDKALOPHIGMSVLDATADEBSLEAIDVNPVY
LRAVEVALTJALMSWGLEPDAVMGSHSGEVAAHVAALSLDEAVLITICHSKLMRTL
GGKATAMVVELSVEAEAEAIRPWNGKIWLGLNGPRSOVLSGEPVAKAEVALGKR
GVFTRWKMDVPSHTPLVEVIRSELGRIRIARTIPRADIPICSTVGEFLDGALEDA
HGMNLIENPVPFAAAQHLASDHGTFEVSPHVLPAVEQTMAGNTEGVALVLSR
RNEDEHRTLSLIGALYSGGRVAMNRYVPGKQVLPPTYPMQREHWMPEAVLAGR
ASRTSSGHPLIGERIDLSAGEVSRIMQVOLGPKHPLIAEHROCVAPALPAAAMD
ALSAKOTLGEFVSLEDFVSALISPEDEERYQLMLSTERPGASATPLFSAPVRE
TPSDELFWALHABGSLKRDASVVEKAPVALDELARCGEASVEAHVOTLKEIGLE
YGPSFVRVRELIRAKGALGRVQLEGLTAEEERYSLHAPLLDCCFQVLAALLEGDS
EGGVYVVALSRVLVLRPAALSHVIRIVSAGASNTITECDVTLIDOGRVAAEVR
RLVCKRTAAASAPADEVDWVYSAQNEPOORAEFGSSRTAGRMVLPADARIGOS
LASGLARAGGVVVEAGPSYERLEADRYINPSKLEDPQALLKEAFSDOSCKGMVH
LYGIDVANSNTEPSEFRALISSTSVYVLIQANQGRMDTPRLMLYTTGVOYIAGS
EQPESPAAPLVGLAAVITHEHPELRCAHYDLSQSPSEETASLTFEELDGGADRIA
LRGAVRVARLGRGISGRKAAEPREFADGTYLITGGLGGLGAEVAKWFIORGARNLA
LTGCAPESEAKVGLIELRANGAEVRYRADVTSADVARLIGSIDSELPRLRGVHA
AGVYADGVLTQIDPEPLRAVLAIPVQGMNLIHVOTGRPLELTFVMSAASMSGPGQ

Query Match 47.9%; Score 48.4; DB 1; Length 49736;
Best Local Similarity 68.4%; Pred. No. 0.071;
Matches 67; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 2 TCGACGTCGTGCGGCGAGCCGACGCTCGCAGAGGTGCTGCGATCCTTGCGCA 61
|||||
DB 45824 TCGACGTCGTGCTGCGGCGAGCCGACGCTCGCAGAGGTGCTGCGATCCTTGCGCGC 45765
|||||
OY 62 AAGGGGTGTCGTGCGGCGAGCTCAAGTGAGCGTCGC 99
DB 45764 CCGGAGTGTCTGCGCGCGGTGATATCAACGTCGC 45727
|||||

Search completed: November 6, 2002, 13:22:46
Job time : 592.625 secs

GNVAAANAFMDVLAAHRRAGLPGLSYNMGMFADVGLAAVATNNGDRALRGVNMMP
AQALEAIGRLVQAOAVVYPLDRQWREPTYSAAOSPILSYMOKKYSTPASKGS
ARELLAAABPQRKGTLEGRLEDRGLIRNEPSRIEDPOFGDGLGDSLGLERBR
IESGLKFLSHTLIYAVPYOALTYIYKLEPYLEAEARTVAPPTOTEPYVSVGAP

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 1197.62 Seconds
(without alignments)
1365.823 Million cell updates/sec

Title: US-09-724-876-2_COPY_32070_32170

Sequence: 1 agccgcgtcgcagcgcctgcg.....tgcgcacatgctctcctca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estc2:*
11: gb_hic:*
12: gb_estl3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inu:*
20: em_gss_pin:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36.8	36.4	516	9	A1532314	
2	34.2	33.9	988	14	B0923465	AGENCOURT
3	33.6	33.3	1200	14	BM804715	AGENCOURT
4	33	32.7	819	12	BG368755	HVSMEL1002
5	32.8	32.5	607	12	BF482518	WHE1795_E
6	32	31.7	714	12	BG305319	fm11901.x

7	32	31.7	1086	13	BM480366	AGENCOURT
8	32	31.7	1500	17	AF057091	AF057091
9	31.8	31.5	650	14	B0241000	TAF05010C
10	31.6	31.3	369	9	AI861213	603011D02
11	31.6	31.3	495	9	AI947509	603022E10
12	31.6	31.3	561	9	AI861177	603017D09
13	31.6	31.3	1856	11	AY109418	Zea mays
14	31.4	31.1	223	13	BI004847	MR4-HN005
15	31.4	31.1	410	12	BG556725	EM1_40_G0
16	31.4	31.1	451	12	BG323133	EM1_15_C0
17	31.4	31.1	559	12	BG464001	EM1_52_E0
18	31.4	31.1	829	9	AL540275	AL540275
19	31.4	31.1	1035	14	B0929758	AGENCOURT
20	31.2	31.1	1453	14	BM926122	AGENCOURT
21	31.2	30.9	1453	14	BM807328	AGENCOURT
22	30.8	30.5	148	17	BH018258	L1394K_0
23	30.8	30.5	639	17	B83523	RPC11.1-1512
24	30.8	30.5	749	12	BG786748	SEAUMC006
25	30.8	30.5	1077	17	AG144203	Pan trogl
26	30.6	30.3	500	9	AL505881	AL505881
27	30.6	30.3	513	10	BE559417	HV_CEB002
28	30.6	30.3	576	14	BQ486420	3524_1_31
29	30.6	30.3	808	17	AQ893261	HS_4832_A
30	30.6	30.3	830	13	BI946866	HVSMEL1000
31	30.6	30.3	871	12	BF264541	HV_CEA000
32	30.4	30.1	585	9	AL821950	AI821950
33	30.4	30.1	585	13	BM137485	WHE0475_F
34	30.4	30.1	796	12	BF628967	HVSMEL000
35	30.4	30.1	1127	12	BG818398	602779940
36	30.2	29.9	240	10	BE419553	WMS014_E1
37	30.2	29.9	347	9	AI183900	Q623409_x
38	30.2	29.9	392	10	BE500283	WHE0981_E
39	30.2	29.9	482	10	BE235903	143528_MA
40	30.2	29.9	494	10	AV621180	AV621180
41	30.2	29.9	498	13	BI344331	372777_MA
42	30.2	29.9	519	14	BQ245561	TAE15023B
43	30.2	29.9	529	10	BE499548	WHE0961_B
44	30.2	29.9	540	9	AA148294	z045f09_s
45	30.2	29.9	572	17	CNS02D20	AI191889 Tetraodon

ALIGNMENTS

RESULT 1
A1532314 516 bp mRNA linear EST 19-APR-2001
LOCUS SD03777.5primg SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION POT2 Drosophila melanogaster cDNA clone SD03777 5prime, mRNA
sequence.

ACCESSION A1532314
VERSION A1532314.1 GI:4446449

KEYWORDS
SOURCE
ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 516)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G. M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

TITLE

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798

JOURNAL

hlt genomic sequence AC005554
Plate: 37 row: G column: 5
High quality sequence stop: 514.
Location/Qualifiers

FEATURES

RESULT 4
 BG368755
 LOCUS
 DEFINITION
 HVSME10020107f Hordeum vulgare 20 DAP spike EST library HVCDNA0010
 (20 DAP) Hordeum vulgare cDNA clone HVSME10020107f, mRNA sequence.
 ACCESSION
 BG368755
 VERSION
 BG368755.1 GI:13257856
 KEYWORDS
 EST.
 SOURCE
 Hordeum vulgare.
 ORGANISM
 Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 1 (bases 1 to 819)
 REFERENCE
 Wing, R., Close, T.J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu
 X., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
 R.D., Close, S.J., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex 20 DAP spike cDNA library
 Unpublished (2001)
 JOURNAL
 COMMENT
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 546
 Seq primer: AATTACCTCCTCAATAAGG
 High quality sequence stop: 717.
 FEATURES
 source
 1..819
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSME10020107f"
 /clone_1lb="Hordeum vulgare 20 DAP spike EST library
 HVCDNA0010 (20 DAP)"
 /tissue_type="20 DAP spike"
 /lab_host="SOLR"
 /note="Vector: lambdaZAP, Site_1: EcoRI, Site_2: XhoI;
 plants were grown in the greenhouse at the University of
 California, Riverside (Fenton, SJ Close, TJ Close). Whole
 spikes with awns trimmed were collected at 20 DAP (Fenton
). Total RNA was prepared, poly(A) RNA was purified, one
 primary unamplified cDNA library was made, and 1 million
 pfu were in vivo excised to give pBluescript SK(-) cDNA
 phagemids in the TJ Close lab at the University of
 California, Riverside (Choi). Phagemids were plated and
 picked at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
 , Rambo, Main). The sequence has been trimmed to remove
 vector sequence and contains a minimum of 100 bases of
 phred value 20 or above. For more details on library
 preparation and sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders Also
 see Close TJ, Wing R, Kleinholz A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley genetics Newsletter 31:25-30.
 (http://wheat.pw.usda.gov/gnpages/bgn/31/cover.html)".
 BASE COUNT
 128 a 255 c 308 g 128 t
 ORIGIN
 Query Match 32.7%; Score 33; DB 12; Length 819;
 Best Local Similarity 65.8%; Pred. No. 60;
 Matches 48; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 GY 2 GCGGCTGCGAGCGCTGCGCGAGCGAGCGAGCGCTTTCGCGCCGACCGGAGC 61
 Db 414 GCGGCGGCGCGAGCGCTGCGCGAGCGAGCGAGCGCTGATGCGCGACGCGGAGCGG 473

GY 62 GCGTGGGCTGGAG 74
 ||||| |
 Db 474 GCGTGGCAGCAGAG 486
 RESULT 5
 BF482518/c
 LOCUS
 DEFINITION
 WHE1795_E07_113S wheat pre-anthesis spike cDNA library Triticum
 aestivum cDNA clone WHE1795_E07_113, mRNA sequence.
 ACCESSION
 BF482518
 VERSION
 BF482518.1 GI:11565819
 KEYWORDS
 EST.
 SOURCE
 bread wheat.
 ORGANISM
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Triticum.
 1 (bases 1 to 607)
 REFERENCE
 Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
 P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.L. and Tong, J.C.
 The structure and function of the expressed portion of the wheat
 genomes - Pre-anthesis spike cDNA library
 Unpublished (2000)
 JOURNAL
 COMMENT
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.
 FEATURES
 source
 1..607
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1795_E07_113"
 /clone_1lb="Wheat pre-anthesis spike cDNA library"
 /tissue_type="Spike before anthesis"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI, Site_2: XhoI; Plants were grown in the
 greenhouse. Whole spike with awns trimmed, white, green
 and yellow anther were collected and total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the TJ Close lab (Choi, Close, Fenton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)".
 BASE COUNT
 108 a 196 c 128 g 175 t
 ORIGIN
 Query Match 32.5%; Score 32.8; DB 12; Length 607;
 Best Local Similarity 64.5%; Pred. No. 65;
 Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 GY 15 CCGGCTCCGAGCGAGCGTTCGCGCGCGCGCGCGCGCGTGGGCTGGAG 74
 Db 216 CCGGCG 157
 GY 75 CGAGGCTGCGCGCATG 90
 Db 156 CGTGGCTGCGCGAAG 141
 RESULT 6

LOCUS	BG305319		714 bp	mRNA	linear	EST 13-FEB-2002
DEFINITION	fml1g01.x1 zebrafish adult retina cDNA Danio rerio CDNA clone 4146216 3 similar to TR:Q9PUS1 Q9PUS1 PROTEASOME SUBUNIT BETA 7 ; , mRNA sequence.					
ACCESSION	BG305319					
VERSION	BG305319.1		GI:13102846			
KEYWORDS	EST.					
SOURCE	zebrafish.					
ORGANISM	Danio rerio					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio. 1 (bases 1 to 714)					
AUTHORS	Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy, 'S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood ,K., Stepec,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Konh.S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.					
TITLE	Washu Zedrafish EST Project 1998					
JOURNAL	Unpublished (1998)					
COMMENT	Contact: Stephen L. Johnson Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: zdrafi@wustl.edu Library constructed by: Chanda Tucker and Gregory Niemi DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: ResourcenzentrumPrimatardatenbank, Berlin, Germany (web address: www.rzpd.de) Seq primer: T7 from Gldco High quality sequence stop: 458. Location/Qualifiers 1..714 /organism="Danio rerio" /strain="wild-type" /db_xref="taxon:7955" /clone="4146216" /clone_lib="zebrafish adult retina cDNA" /sex="mixed" /dev_stage="1-2 years" /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')" /note="Vector: Lambda Zap II (pBluescript SK-); Site_1: ECORI; Site_2: SalI; This zebrafish library was constructed by Dr. Susan E. Brockerhoff (email: sbrockre@u.washington.edu) RZPD library number: 760"					
BASE COUNT	170 a 182 c 185 g 177 t					
ORIGIN						
Query Match	31.7%; Score 32; DB 12; Length 714;					
Best Local Similarity	60.2%; Pred. No. 1.1e+02;					
Matches 53; Conservative	0; Mismatches 35; Indels 0; Gaps 0;					
OY	14	GCTGGTCCCGAGCAGCGTGCAAGACTCTTCGCGCCGACCGCAGCGCGGTGGGTGA	73			
Db	602	GGCAGACGCCGGGGGAGAGAATGAACCTCCACAGTTGAGAGAGATGACCTGTGGGCA	661			
OY	74	GCGAAGGCTGGCGCATGCTCTCGCTCAA	101			
Db	662	TCTAAGTGTCCGCGAAGTACCTGTCAA	689			
RESULT 7						
LOCUS	BM480366		1086 bp	mRNA	linear	EST 05-FEB-2002
DEFINITION	AGENCOURT.6468237 NIH_MGC_88 Homo sapiens CDNA IMAGE:5574241 5', mRNA sequence.					
ACCESSION	BM480366					
VERSION	BM480366.1		GI:18529408			
KEYWORDS	EST.					
SOURCE	human.					

	ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. REFERENCE AUTHORS NIH-MGC http://mgc.nci.nih.gov/. TITLE National Institutes of Health, Mammalian Gene Collection (MGC) JOURNML Unpublished (1999)
COMMENT	Contact: Robert Strusberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CNA Library Preparation: Life Technologies, Inc. cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LRAM1232 row: j column: 02 High quality sequence stop: 322.	
FEATURES	Location/Qualifiers 1..1086 /organism="Homo sapiens" /db_xref="taxon:9606" /collection="IMAGE:5574241" /clone_id="NIH_MGC_88" /tissue_type="tumor adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: small intestine; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; size=47 primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH/MGC Library."	
BASE COUNT	81 A 266 C 268 G 65 T 406 others	
ORIGIN		
Query Match	31.7%; Score 32; DB 13; Length 1086;	
Best Local Similarity	57.1%; Pred. NO. 1.le+02;	
Matches	56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;	
OY	2 GCCGGCTCGAGACGCTGCCGAGCACGTTCGCAAGACTTGCGCGCAGCCGACG 61 Db 954 GC CGCGGGCGAGAGCGCGCGCGGGGTCCGCGCGCGCCNCNGGGGGCGGCGCG 1013	
OY	62 GCGTGGGTGAGAGCGAAGGCTGCCCATGCTCTCTCTC 99 Db 1014 GTATGGGGGGGCGAGCGCGCGCGCCGACTCCCC 1051	
RESULT 8		
APOF57091/C	AF057091 1500 bp DNA linear GSS 21-FEB-2001	
LOCUS	AP057091 Human Homo sapiens genomic clone 25f probe, DNA sequence.	
DEFINITION	AP057091	
ACCESSION	AF057091	
VERSION	AF057091.1 GI:3142398	
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	Janer,M. and Geraghty,D.E. The human major histocompatibility complex: 42,221 bp of genomic sequence, high-density sequence-tagged site map, evolution, and polymorphism for HLA class I	
TITLE	Genomics 51 (1), 35-44 (1998)	
JOURNAL MEDLINE	98360090	
COMMENT	Contact: Janer M Geraghty Lab Fred Hutchinson Cancer Research Center 1100 Fairview Ave., D2-100, Seattle, WA 98109-1024, USA Email: geraghtye@fhcrc.org Class: unknown.	
FEATURES	Location/Qualifiers 1..1500	
SOURCE		

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p21.3: HLA Class I region"
/clone="25R probe"
/clone_lib="Human"
/Note="Vector: pMOS Blue"

BASE COUNT      269 a  450 c  406 g  375 t
ORIGIN

Query Match      31.7%; Score 32; DB 17; Length 1500;
Best Local Similarity 58.3%; Pred. No. 1.2e+02;
Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY      2  GCGGCTCCGAGCGCTGCTCCGAGCGAGCGTGCAGAGCTTCTCGCCGCGACCGAGC 61
          |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db      630 GCGGCTGCTAGTGCCTCCGCGCCCTCGAGAGCTGCGAGGCTGCTGGAGAGGTAGACA 571
          |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
OY      62  GCGTGGGTGAGCGAAGCGCTGCGCCATGCTCTGTC 97
          |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db      570 GGCATGGGAGAGTGAAGTGGGGGCTGACCCAGC 535

RESULT 9
BQ241000/c      650 bp  mRNA  linear  EST 03-MAY-2002
LOCUS      TAE05010C09R TAE05 Triticum aestivum cDNA clone TAE05010C09R, mRNA
DEFINITION      sequence.
ACCESSION      BQ241000
VERSION      BQ241000.1 GI:20436876
KEYWORDS      bread wheat.
SOURCE      Triticum aestivum
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
              Triticaceae; Triticum.
              1 (bases 1 to 650)
REFERENCE      1 Cloutier,S.
              Wheat functional genomics - glenlea developing seeds cDNA libraries
              Unpublished (2002)
AUTHORS      Contact: Dr. Sylvie Cloutier
              Cereal Research Centre, Agriculture and Agri-Food Canada
              195 D'afce Rd, Winnipeg, MB, Canada R3T 2M9
              Tel: (204) 983-2340
              Fax: (204) 983-4604
              Email: scloutier@agr.ca
              was cloned directionally, not all sequences generated with reverse
              primer were from the 5' end (same with forward primer and 3' end).
              Average insert size is >2.0 kb
              Plate: 010 row: C column: 09
              Seq primer: M13 Reverse.
FEATURES
    source
        location/Qualifiers
            1..650
                /organism="Triticum aestivum"
                /cultivar="Glenlea"
                /db_xref="taxon:4565"
                /clone="TAE05010C09R"
                /clone_lib="rap05"
                /tissue_type="developing seeds"
                /dev_stage="5 days after anthesis"
                /lab_host="E. coli DH10B"
                /note="Vector: pSPORT-P (Invitrogen Technologies); Site_1:
                NotI; Site_2: MluI; mRNA obtained from wheat seeds of
                cultivar Glenlea 5 days post-anthesis"
BASE COUNT      111 a  235 c  162 g  141 t  1 others
ORIGIN

Query Match      31.5%; Score 31.8; DB 14; Length 650;
Best Local Similarity 64.0%; Pred. No. 1.2e+02;
Matches 48; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY      16  CTGGCTCCCGAGCGAGCGTCTTCGCGCCGCGAGCGCGGTGGGTGAGC 75
          |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db      277 CTGGCGCCGACCGCGGAGCGGAGCGGCGCGGCGAAGCGCGCGCGCGAGC 218

```

```

OY      76  GAAGCTGCGCGCANG 90
          |  |||||  |||
Db      217 GCGGCTGCGCGANG 203

RESULT 10
A1861213/c      369 bp  mRNA  linear  EST 19-JUL-1999
LOCUS      603011D02.x1 603 - stressed root cDNA library from Wang/Bohnert lab
DEFINITION      Zea mays cDNA, mRNA sequence.
ACCESSION      A1861213
VERSION      A1861213.1 GI:5525215
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays.
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoidae; Andropogoneae; Zea.
              1 (bases 1 to 369)
REFERENCE      Walbot,V
              Maize ESTs from various cDNA libraries sequenced at Stanford
              University
              Unpublished (1999)
              Contact: Walbot V
              Department of Biological Sciences
              Stanford University
              855 California Ave, Palo Alto, CA 94304, USA
              Tel: 650 723 2227
              Fax: 650 725 8221
              Email: walbot@stanford.edu
              Plate: 603011 row: D column: 02.
FEATURES
    source
        location/Qualifiers
            1..369
                /organism="Zea mays"
                /cultivar="B73"
                /db_xref="taxon:4577"
                /clone_lib="603 - stressed root cDNA library from
                Wang/Bohnert lab"
                /tissue_type="seedling"
                /dev_stage="salt stress"
                /lab_host="E. coli XL Gold"
                /note="Organ: root; Vector: pBluescriptII SK(+); XR;
                Seedling stressed root cDNA library from Wang/Bohnert lab"
BASE COUNT      74 a  127 c  99 g  69 t
ORIGIN

Query Match      31.3%; Score 31.6; DB 9; Length 369;
Best Local Similarity 58.5%; Pred. No. 1.2e+02;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY      8  TGCAGGCGCTGCTCCGAGCGAGCGTGCAGAGCTTCTCGCCGCGACCGCGCTGC 67
          |||  |||||  |  |||  |||||  |||  |||||  |||  |||||
Db      295 TGGCTGCTCTCCAGAGCGACGACGAGCGGCGGCGGCGACATCTTCGTGG 236
          |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
OY      68  GGTGAGCGAAGCGCGCGCATGCTGCTGCA 101
          |  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db      235 GCGGCGCGTGGCGAGCATCGACCTGGCCGA 202
          |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||

RESULT 11
A1947509/c      495 bp  mRNA  linear  EST 19-AUG-1999
LOCUS      603022E10.x1 603 - stressed root cDNA library from Wang/Bohnert lab
DEFINITION      Zea mays cDNA, mRNA sequence.
ACCESSION      A1947509
VERSION      A1947509.1 GI:5739819
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays.
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoidae; Andropogoneae; Zea.

```



```

RESULT 14
BI004847      223 bp  mRNA      linear  EST 13-JUN-2001
LOCUS         BI004847
DEFINITION    NM4-HN0056-150301-001-g10 HN0056 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BI004847
VERSION       BI004847.1  GI:14408921
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 223)
AUTHORS       Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE         Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE       20202663
COMMENT       Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR4&t2=MR4-HN0056-
150301-001-g10&t3=2001-03-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 38
High quality sequence stop: 223.
FEATURES
source
location/Qualifiers
1..223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="HN0056"
/dev_stage="Adult"
/note="Organ: head,normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT    41 a 72 c 30 t 1 others
ORIGIN
Query Match      31.1%; Score 31.4; DB 13; Length 223;
Best Local Similarity 59.6%; Pred. No. 1.3e+02;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 2 GCCGGTCGAGAGCCTGCTCCGACGAGACGCTTCGCGCCGACGCGACG 61
|| || || || || || || || || || || || || || || || || ||
DB 85 GCACGGTCAGAGGTGGCGCCGCGTGCATTCGACGACCTTCGGGCGCCGTCCTGCG 144
|| || || || || || || || || || || || || || || || || ||
QY 62 GCGTGGGTGGAGCGAAGCTGCGCCATG 90
|| || || || || || || || || || || || || || || || || ||
DB 145 CCGCCAGGCTGACCGAATGCTGCGCGGTG 173
|| || || || || || || || || || || || || || || || || ||

```

```

RESULT 15
BG556725      410 bp  mRNA      linear  EST 10-APR-2001
LOCUS         BG556725
DEFINITION    EML40_G08.g1_A002 Embryo 1 (EML) sorghum bicolor cDNA, mRNA
sequence.

```

```

ACCESSION     BG556725
VERSION       BG556725.1  GI:13585723
KEYWORDS
SOURCE        Sorghum.
ORGANISM      Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
ciade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE     1 (bases 1 to 410)
AUTHORS       Reid,S.P., Cordomier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE         An EST database from Sorghum: developing embryos
JOURNAL       Unpublished (2000)
COMMENT       Contact: Cordomier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: polyTmix
High quality sequence start: 46
High quality sequence stop: 405
POLYA=yes.
FEATURES
source
location/Qualifiers
1..410
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_id="Embryo 1 (EML)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT    69 a 120 c 164 g 57 t
ORIGIN
Query Match      31.1%; Score 31.4; DB 12; Length 410;
Best Local Similarity 59.6%; Pred. No. 1.4e+02;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 8 TTGAGAGGCTGCGTCCGACGAGCGTGTGCGCGCGACGCGCGCGG 67
|| || || || || || || || || || || || || || || || || ||
DB 197 TTGCTGCTTACCAAGACAGCGACGCGCATGTGAGGCCGCGACATCTTCTGTG 256
|| || || || || || || || || || || || || || || || || ||
QY 68 GTGAGCGAAGGCTGCGCCATGCTCTG 96
|| || || || || || || || || || || || || || || || || ||
DB 257 GCGGCGGCTGCGCAGCGACATCGCACCTG 285
|| || || || || || || || || || || || || || || || || ||

```

Search completed: November 6, 2002, 15:52:31
Job time : 1201.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 ; Search time 27.125 Seconds
(Without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_32070_32170

Perfect score: 101

Sequence: 1 agccgctgcgagcctgcgc.....tgccatctctctctca 101

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 310279 segs, 16577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.4	53.9	15872	10	US-09-861-289-1
2	51.4	50.9	4041	10	US-09-861-289-36
3	51.4	50.9	11220	10	US-09-861-289-32
4	51.4	50.9	36778	10	US-09-861-289-5
5	45	44.6	13842	10	US-09-861-289-30
6	40.4	40.0	4689	10	US-09-861-289-34
7	32.2	31.9	643	10	US-09-924-256A-53
8	29	28.7	643	10	US-09-924-256A-31
9	28.6	28.3	4364	10	US-09-985-675-9
10	28.4	28.1	2242	10	US-09-924-859A-2
11	28.4	28.1	3611	10	US-09-923-490-1
12	28	27.7	1395	10	US-09-815-242-4093
13	28	27.7	2410	10	US-09-006-298-26
14	28	27.7	4180	10	US-09-880-107-3714
15	27.6	27.3	2162	10	US-09-799-875-10
16	27.6	27.3	2454	10	US-09-757-982-7
17	27.4	27.1	765	10	US-09-924-256A-57
18	27.2	26.9	357	10	US-09-777-564-881
19	27.2	26.9	368	10	US-09-777-564-553

20	27.2	26.9	412	10	US-09-960-352-5578	Sequence 5578, Ap
21	27.2	26.9	432	10	US-09-960-352-11709	Sequence 11709, A
22	27.2	26.9	555	10	US-09-919-580-453	Sequence 453, App
23	27.2	26.9	558	10	US-09-920-300A-5	Sequence 5, Appl
24	27.2	26.9	538	12	US-10-033-528-5	Sequence 5, Appl
25	27.2	26.9	589	10	US-09-864-761-19479	Sequence 19479, A
26	27.2	26.9	672	10	US-09-880-107-3737	Sequence 3737, Ap
27	27.2	26.9	960	10	US-09-864-761-2763	Sequence 2763, Ap
28	27	26.7	2553	10	US-09-815-242-7690	Sequence 7690, Ap
29	27	26.7	4826	10	US-09-772-304A-1	Sequence 1, Appl
30	26.8	26.5	429	10	US-09-884-441-309	Sequence 309, App
31	26.8	26.5	719	10	US-09-923-779-1	Sequence 1, Appl
32	26.6	26.3	888	10	US-09-815-242-7873	Sequence 7873, Ap
33	26.6	26.3	3035	10	US-09-864-864-311	Sequence 311, App
34	26.4	26.1	1269	10	US-09-764-847-1188	Sequence 397, App
35	26.4	26.1	1269	10	US-09-885-478-1	Sequence 1, Appl
36	26.4	26.1	1385	10	US-09-925-776-1	Sequence 1, Appl
37	26.4	26.1	1655	10	US-09-925-299-205	Sequence 105, App
38	26.4	26.1	3884	10	US-09-820-809-14	Sequence 14, Appl
39	26.2	25.9	261	10	US-09-728-445-393	Sequence 1188, Ap
40	26.2	25.9	484	10	US-09-814-292-46	Sequence 393, App
41	26.2	25.9	575	10	US-09-823-847-5	Sequence 46, Appl
42	26.2	25.9	1680	10	US-09-823-847-5	Sequence 5, Appl
43	26	25.7	198	10	US-09-895-686-55	Sequence 55, Appl
44	26	25.7	266	10	US-09-923-876-819	Sequence 819, App
45	26	25.7	272	10	US-09-923-876-2809	Sequence 2809, Ap

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OR INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-1

Query Match 53.9%; Score 54.4; DB 10; Length 15872;
Best Local Similarity 72.9%; Pred. No. 2.9e-07;
Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CGGCTGCGAGCGCTGCTCCGACGAGCGTGACGACTTCTCGGCGCGACGCGACGG 62
DB 9136 CCGGACAGAGGCGCTGCGCGCGACGCGCGCTTCGCGCGCGCGCGCGACGG 9195

QY 63 CGTGGGTGAGCGGACGAGCGCTGCGCGCGCTCTCTCTCT 98
DB 9196 CACCGCTGCGCGCGAGGCGCTGCGCGCTCTCTCTCT 9231

RESULT 2
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```

; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-36
```

```
Query Match          50.9%; Score 51.4; DB 10; Length 4041;
Best Local Similarity 69.3%; Pred. No. 1.9e-06;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 AGCCGCTCGAGAGCCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGACGCGAC 60
DB 739 AGCAGCCACAGAGCCATGCGCCCGACGCGCCGACCAAGGCGCTTCGCGCGCGAC 798
QY 61 GCGGTGGGTGAGCGAGCGCTGCCCATGCTCTCTCTCAA 101
DB 799 GCGCTCGCTGGGCGAGGCGCTGCGCGTACTCTCTCTCGA 839
```

```

RESULT 3
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32
```

```
Query Match          50.9%; Score 51.4; DB 10; Length 11220;
Best Local Similarity 69.3%; Pred. No. 2e-06;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 AGCCGCTCGAGAGCCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGACGCGAC 60
DB 742 AGCCGCTACAGCGGCGGTGCGCGCGACGCGCGCTTCGATCTCTCTGAGAGGCGAC 801
QY 61 GCGGTGGGTGAGCGAGCGCTGCCCATGCTCTCTCTCAA 101
DB 802 GCGACACAGCTGTGTCGAGGCGCTGCGCGTCTCTCTCTCGA 842
```

```

RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
```

```

; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match          50.9%; Score 51.4; DB 10; Length 36778;
Best Local Similarity 69.3%; Pred. No. 2.1e-06;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 1 AGCCGCTCGAGAGCCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGACGCGAC 60
DB 16429 AGCGTACAGCGCGGCTGCGCGCGACGCGCGGTGCAAGTCTCTCTCTGAGAGCGAC 16488
QY 61 GCGGTGGGTGAGCGAGCGCTGCCCATGCTCTCTCTCAA 101
DB 16489 GCGACACAGCTGTGTCGAGGCGCTGCGCGTCTCTCTCTCGA 16529
```

```

RESULT 5
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30
```

```
Query Match          44.6%; Score 45; DB 10; Length 13842;
Best Local Similarity 65.3%; Pred. No. 0.00013;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
QY 1 AGCCGCTCGAGAGCCTGCTCCCGACGAGCGGTGCAAGAGCTTCTCGGCCGACGCGAC 60
DB 3961 AGCCGCTACAGCGGCGGTGCGCGCGACGCGCGGTGCAAGAGCTTCTCTCTGAGAGGCGAC 4020
QY 61 GCGGTGGGTGAGCGAGCGCTGCCCATGCTCTCTCTCAA 101
DB 4021 GCGACACAGCTGTGTCGAGGCGCTGCGCGTCTCTCTCTCGA 4061
```

```

RESULT 6
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
```


US-09-985-675-9

Query Match	28.38;	Score	28.6;	DB	10;	Length	4364;
Best Local Similarity	59.08;	Pred. No.	5.2;				
Matches	49;	Conservative	0;	Mismatches	34;	Indels	0;
				Gaps			0;

QY	18	GGCTCCGACGACGGGACAGACCTTCTGGCCGCGACGACGGGCTGGAGGACCGA	77
Db	744	GGACTCCGCGCAGACAGACCCGACCGCGGGTGGGCGCAGCGCGCAGCGCGCGAGACC	685
QY	78	AGCTGCGCCATGCTCCTGCTCA	100
Db	684	CGGCGGCGCCATGCTCTCGGCTCA	662

RESULT 10
US-09-924-859A-2/c
; Sequence 2, Application US/09924859A
; Date of Invention 09/09/2009

Query Match	28.1%	Score 28.4	DB 10	Length 2742
Best Local Similarly	60.3%	Pred. No. 5.8		
Matches 47	Conservative 0	Mismatches 31	Indels 0	Gaps 0

Qy 23 CCGACGAGACGCTCCAGACCTTCTGGGCGCGACGCGAGCGCTGGGTGGAGCGAAGGCT 82
Db 85 CCGCGACBACBACBACCCGACCCCGGTGGCGCGCGACGCGTCAAGCGGCGGACCCCGGCC 28
Qy 83 GCGCGATGCTCTGCTCA 100
Db 25 GCGCGATGCTCCGCTCA 8

RESULT 11
 US-09-223-490-1/c
 Sequence 1, Application US/09223490
 Patent No. US20020117325A1
 GENERAL INFORMATION:
 APPLICANT: Godowski, Paul J.
 APPLICANT: Mark, Melanie R.
 APPLICANT: Scadden, David T.
 APPLICANT: Baker, Kevin P.
 APPLICANT: Barton, Will F.
 TITLE OF INVENTION: Protein Tyrosine Kinases
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
 COMPUTER: IBM PC compatible

```

:
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/223,490
: FILING DATE:
:

```

```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: 08/170,558
3 FILING DATE:
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Hasak, Janet E.
6 REGISTRATION NUMBER: 28,616
7 REFERENCE/DOCKET NUMBER: 854C1
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 415/225-1896
10 TELEFAX: 415/952-9881
11 TELEX: 910/371-7168
12 INFORMATION FOR SEQ ID NO: 1:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 3611 bases
15 TYPE: nucleic acid
16 STRANDEDNESS: single
17 TOPOLOGY: linear
18
19 OS-09-223-490-1

```

Query Match	28.1%	Score 28.4	DB 10	Length 3611
Best Local Similarity	60.3%	Pred. No. 5.8		
Matches 47; Conservative	0	Mismatches 31	Indels 0	Gaps 0

OY	23	CCGAGCAGCGTTCACAGACTTCTGGGCCGACGCAGCGGTGGGTGAGCAAGGCT	83
Dd	91	CCGCGACAGCACAGCCCCGAGCCCGGGGGGGGGGAGCGGACGAGCGGGGAGCCCGCGCC	323
OY	83	GCGCGCATGCTCTGCTCA	100
Dd	31	GCCCATGCTCTGCTCA	14

```

RESULT 12
US-09-815-242-4093
Sequence 4093, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlser, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Treawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4093

```

LENGTH: 1395
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4093

Query Match 27.7%; Score 28; DB 10; Length 1395;
Best Local Similarity 55.0%; Pred. No. 7.3;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2 GCCGCTGCGAGCGCTGCTCCGACGAGCGTGCAGAGCTTCTCGCCGACCGCAGC 61
DB 482 GCCGCGAACCGAACCCGACCGACCGCGCGGCGGATCGCCGAGCATCTCCGACG 541
QY 62 GCGTGGGTGGAGCGAGCGTGGCGCATGCTCTGCTCAA 101
DB 542 GCGTGTGCTTGGCGAATGGCAGCGCATGCTCGGATCAA 581

RESULT 13
US-09-006-298-26
Sequence 26, Application US/09006298
Patent No. US2002008224A1
GENERAL INFORMATION:
APPLICANT: Jolly, Douglas J.
APPLICANT: Moore, Margaret D.
APPLICANT: Chada, Sunil
TITLE OF INVENTION: NON-IMMUNOGENIC PRODRUGS AND SELECTABLE
MARKERS FOR USE IN GENE THERAPY
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,298
FILING DATE: 13-JAN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 930049, 459
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 2410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 17..2398
US-09-006-298-26

Query Match 27.7%; Score 28; DB 10; Length 2410;
Best Local Similarity 56.5%; Pred. No. 7.4;
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 5 GCGTCGAGCGCTGCTCCGACGAGCGTTCGCGCGACCGCAGCGCG 64
DB 1746 GCGTCCCGTACCTCCAGAAAGAGTGCTCAAGACCCCTCAGTCAGGCGCTGTG 1805
QY 65 TGGGTGGAGCGAGGCTGCGCCATGCTCTG 96

DB 1806 TGTGTGCGAGGAGGCTTCTCCTGCACAG 1837

RESULT 14
US-09-880-107-3714
Sequence 3714, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3714
LENGTH: 4180
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X17094
US-09-880-107-3714

Query Match 27.7%; Score 28; DB 10; Length 4180;
Best Local Similarity 56.5%; Pred. No. 7.6;
Matches 52; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 5 GCGTCGAGCGCTGCTCCGACGAGCGTGCAGAGCTTCTCGCCGACCGCAGCGCG 64
DB 1946 GCGTCCCGTACCTCCAGAAAGAGTGCTCAAGACCCCTCAGTCAGGCGCTGTG 2005
QY 65 TGGGTGGAGCGAGGCTGCGCCATGCTCTG 96
DB 2006 TGTGTGCGAGGAGGCTTCTCCTGCACAG 2037

RESULT 15
US-09-799-875-10
Sequence 10, Application US/09799875
Patent No. US20020034780A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel
APPLICANT: Kapeller-Libermann, Rosana
TITLE OF INVENTION: No. US20020034780A1e1 Human Protein Kinases and Uses
FILE REFERENCE: 35800/209996
CURRENT APPLICATION NUMBER: US/09/799,875
CURRENT FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: 60/182,059
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: 09/659,287
PRIOR FILING DATE: 2000-09-12
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 2162
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (76)...(1818)
US-09-799-875-10

Query Match 27.3%; Score 27.6; DB 10; Length 2162;
Best Local Similarity 58.5%; Pred. No. 9.6;

	Matches	48;	Conservative	0;	Mismatches	34;	Indels	0;	Gaps	0;
Qy	5	GGCTGGAGGCTTGGCTCCCGAGGAGGTGCAAGAGCTTCTGGCCGACCGAGCGCG	64							
Db	54	GGCAGCGGGGGCGGCGGCCACCATGGGGGTGGCAGAGCGCTGGCGCGCGCCTGCAGCT	113							
Qy	65	TGGGGTGGAGCGAAGGCTGCGC	86							
Db	114	GGGTCGAGCGCTGCTGCGC	135							

Search completed: November 6, 2002, 20:31:24
 Job time : 42.125 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_32070_32170

Perfect score: 101

Sequence: 1 agccgctgcgagcctgcgc.....tgccatctctctgtctca 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 682724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCFUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appl
2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appl
3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appl
4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appl
5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appl
6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appl
7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appl
8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appl
9	56.2	55.6	33529	4	US-09-144-085-3	Sequence 3, Appl
10	54.6	54.1	751	3	US-09-010-809-4	Sequence 4, Appl
11	54.4	53.9	15872	4	US-09-105-537-1	Sequence 1, Appl
12	54.4	53.9	43280	2	US-08-804-227C-1	Sequence 1, Appl
13	53.2	52.7	80161	4	US-09-370-700-1	Sequence 1, Appl
14	53.2	52.7	80161	4	US-09-370-700-1	Sequence 1, Appl
15	52.6	52.1	28958	1	US-08-258-261B-6	Sequence 6, Appl
16	52.6	52.1	28958	1	US-08-457-342-6	Sequence 6, Appl
17	52.6	52.1	28958	1	US-08-457-342-6	Sequence 6, Appl
18	52.6	52.1	28958	1	US-08-457-342-6	Sequence 6, Appl
19	52.6	52.1	28958	1	US-08-458-076A-6	Sequence 6, Appl
20	52.6	52.1	28958	1	US-08-764-233A-4	Sequence 6, Appl
21	52.6	52.1	28958	1	US-08-457-335A-6	Sequence 6, Appl
22	52.6	52.1	28958	1	US-08-729-214-6	Sequence 6, Appl
23	52.6	52.1	28958	3	US-09-028-934-6	Sequence 6, Appl
24	52.6	52.1	49377	1	US-08-764-233A-1	Sequence 1, Appl
25	51.4	50.9	1462	4	US-09-434-288-4	Sequence 4, Appl
26	51.4	50.9	4041	4	US-09-105-537-36	Sequence 36, Appl
27	51.4	50.9	11220	4	US-09-105-537-32	Sequence 32, Appl

28	51.4	50.9	36778	4	US-09-105-537-5	Sequence 5, Appl
29	51.4	50.9	38506	3	US-09-320-878-19	Sequence 19, Appl
30	50.2	49.7	44377	2	US-08-804-227C-7	Sequence 7, Appl
31	50.2	49.7	44377	2	US-08-804-198-1	Sequence 1, Appl
32	50	49.5	11219	1	US-07-642-734C-1	Sequence 1, Appl
33	50	49.5	11219	3	US-08-439-009A-1	Sequence 1, Appl
34	49.8	49.3	20235	1	US-07-642-734C-3	Sequence 3, Appl
35	49.8	49.3	20235	3	US-08-439-009A-3	Sequence 3, Appl
36	47.6	47.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl
37	47.6	47.1	4411529	4	US-09-103-840A-1	Sequence 1, Appl
38	45.8	45.3	13987	2	US-08-804-227C-13	Sequence 13, Appl
39	45.2	44.8	50937	4	US-09-428-517-1	Sequence 1, Appl
40	45	44.6	13842	4	US-09-105-537-30	Sequence 30, Appl
41	43.8	43.4	423	4	US-09-144-085-5	Sequence 5, Appl
42	41.8	41.4	1419	4	US-09-434-288-2	Sequence 2, Appl
43	40.4	40.0	4689	4	US-09-105-537-34	Sequence 34, Appl
44	39.6	39.2	333	4	US-09-060-756-318	Sequence 318, App
45	39.6	39.2	1434	4	US-09-434-288-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match      100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGCTGCGAGGCTGCTCCGACGAGCGTGCAAGAGCTTCTGCGCCGACCGAC 60
    |||||||
Db 37682 AGCCGCTGCGAGGCTGCTCCGACGAGCGTGCAAGAGCTTCTGCGCCGAC 37741

QY 61 GCGGTGGGTGAGCGAAGGCTGCGCATGCTGCTGCTCA 101
    |||||||
Db 37742 GCGGTGGGTGAGCGAAGGCTGCGCATGCTGCTGCTCA 37782

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
```

;; CURRENT FILING DATE: 2000-05-10
;; PRIORITY APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 37741

OY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 37782

RESULT 3
US-09-567-969-1
;; Sequence 1, Application US/09567969
;; Patent No. 6355457
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/567,969
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 37741

OY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 37782

RESULT 4
US-09-568-480-1
;; Sequence 1, Application US/09568480
;; Patent No. 6355458
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,480
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-480-1

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,480
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 37741

OY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 37782

RESULT 5
US-09-568-486-1
;; Sequence 1, Application US/09568486
;; Patent No. 6355459
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,486
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 60
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTCTCGCCGCGAGCCGAC 37741

OY 61 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 101
Db 37742 GCGGTGGGTGAGCGAAGGCTGCGCCATGCTCTCTCA 37782

RESULT 6
US-09-568-472-1
;; Sequence 1, Application US/09568472
;; Patent No. 6358719
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,472
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-472-1

```

: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-472-1

```

```

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 37682 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 37741

```

```

Qy 61 GCGGTGGGGTGAGAGGCTGCGCCATGCTCTCTCA 101
Db 37742 GCGGTGGGGTGAGAGGCTGCGCCATGCTCTCTCA 37782

```

```

RESULT 7
US-09-567-899-1
: Sequence 1, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-899-1

```

```

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 37682 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 37741

```

```

Qy 61 GCGGTGGGGTGAGAGGCTGCGCCATGCTCTCTCA 101
Db 37742 GCGGTGGGGTGAGAGGCTGCGCCATGCTCTCTCA 37782

```

```

RESULT 8
US-09-443-501A-2
: Sequence 2, Application US/09443501A
: Patent No. 6303342
: GENERAL INFORMATION:

```

```

: APPLICANT: Kosan Biosciences, Inc.
: APPLICANT: Julien, Bryan
: APPLICANT: Katz, Leonard
: APPLICANT: Khosla, Chaitan
: APPLICANT: Tang, Li
: APPLICANT: Ziermann, Rainer
: TITLE OF INVENTION: Recombinant Methods and Materials for Producing
: TITLE OF INVENTION: Epothilone and Epothilone Derivatives
: FILE REFERENCE: 30062-20031.00
: CURRENT APPLICATION NUMBER: US/09/443,501A
: CURRENT FILING DATE: 1999-11-19
: PRIOR APPLICATION NUMBER: US 60/130,560
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: US 60/122,620
: PRIOR FILING DATE: 1999-03-03
: PRIOR APPLICATION NUMBER: US 60/119,386
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: US 60/109,401
: PRIOR FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 71989
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

```

```

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60
Db 32070 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 32129

```

```

Qy 61 GCGGTGGGGTGAGAGGCTGCGCCATGCTCTCTCA 101
Db 32130 GCGGTGGGGTGAGAGGCTGCGCCATGCTCTCTCA 32170

```

```

RESULT 9
US-09-144-085-3
: Sequence 3, Application US/09144085
: Patent No. 6280999
: GENERAL INFORMATION:
: APPLICANT: Gustafsson, Claes
: APPLICANT: Betlach, Mary C.
: APPLICANT: Ashley, Gary
: APPLICANT: Julien, Bryan
: APPLICANT: Ziermann, Rainer
: TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
: FILE REFERENCE: 30062-20020.20
: CURRENT APPLICATION NUMBER: US/09/144,085
: CURRENT FILING DATE: 1998-08-31
: EARLIER APPLICATION NUMBER: 09/010,809
: PRIOR FILING DATE: 1998-01-22
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 33529
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-144-085-3

```

```

Query Match 55.6%; Score 56.2; DB 4; Length 33529;
Best Local Similarity 72.3%; Pred. No. 3.7e-07;
Matches 73; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
Qy 1 AGCCGGCTGCAGAGGCTGCTCCGACGAGGTCAGAGCTTCTCGCCGACCGAC 60

```

Db 376 AGCCGGCTCAAGGGGATGCCCGGACGGCCGCTGCAGAGACTTCTTGGCGGAGCTGAC 435
QY 61 GGCGTGGGTGAGAGAGGCTGCGCATGCTCTGCTCAA 101
Db 436 GGCGTCACTGTGTCGAGGGGTGCGGATGCTGTCTGTA 476

RESULT 10

US-09-010-809-4
; Sequence 4, Application US/09010809B
; Patent No. 6090601
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Bellach, Mary C.
; TITLE OF INVENTION: Epithelione Polyketide Synthases and Encoding DNA
; FILE REFERENCE: 30062-20020.00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 751
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-010-809-4

Query Match Best Local Similarity 71.1%; Score 54.6; DB 3; Length 751;
Matches 72; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 AGCCGGCTGAGAGGCTGCTCCGACGAGCGTGCAGAGACTTCTGCGCCGACGCGAC 60
Db 130 AGTCGTGCCCGGGGCTTGGCGAGCGTGGTGCAAGAGCTTCTGCGCAGGCTGAC 169
QY 61 GGCGTGGGTGAGAGGCTGCGCATGCTCTGCTCAA 101
Db 190 GGCGTCACTGTGTCGAGGGGTGCGGATGCTGTCTGTA 230

RESULT 11

US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match Best Local Similarity 72.9%; Score 54.4; DB 4; Length 15872;
Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CGCGCTGCAGAGGCTGCTCCGACGAGGCTGCAGAGCTTCTGCGCCGACGCGACG 62
Db 9136 CGCGCAGAGAGGCTGCTCCGACGAGGCTGCAGAGCTTCTGCGCCGACGCGACG 9195

QY 63 CGTGGGTGAGAGGCTGCGCATGCTCTGCT 98
Db 9196 CACCGGCTGGGCGGAGGCGTGGCTGCTGCT 9231

RESULT 12

US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kustos, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: *20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
US-08-804-227C-1

Query Match Best Local Similarity 72.9%; Score 54.4; DB 2; Length 43280;
Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CGCGCTGCAGAGGCTGCTCCGACGAGGCTGCAGAGCTTCTGCGCCGACGCGACG 62
Db 36998 CGCGCAGAGAGGCTGCTCCGACGAGGCTGCAGAGCTTCTGCGCCGACGCGACG 37057

QY 63 CGTGGGTGAGAGGCTGCGCATGCTCTGCT 98
Db 37058 CACCGGCTGGGCGGAGGCGTGGCTGCTGCT 37093

RESULT 13
US-09-036-987A-1


```

; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-Mar-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

Query Match 52.7%; Score 53.2; DB 3; Length 80161;
Best Local Similarity 71.4%; Pred. No. 2.5e-06;
Matches 70; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 AGCGGCTGGAGGCGCTCCGAGCGGCAAGAGCTTCGCGCCGACGCCGAC 60
Db 60538 AGCGGGAAGCGGCGACCTCGGATGGCGGCAAGCCCTTCGGGATGCCGCGAC 60597
QY 61 GCGGTGGGTGAGAGCGAGGCTGGCCATGCTCTGCT 98
Db 60598 GGCACCGCGTGGGCGGAGGCGCGCGAATGCTGCTGCT 60635

RESULT 14
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1

```

```

; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-1

Query Match 52.7%; Score 53.2; DB 4; Length 80161;
Best Local Similarity 71.4%; Pred. No. 2.5e-06;
Matches 70; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 AGCGGCTGGAGGCGCTCCGAGCGGAGGCTGGCCATGCTCTGCT 98
Db 60538 AGCGGGAAGCGGCGACCTCGGATGGCGGCAAGCCCTTCGGGATGCCGCGAC 60597
QY 61 GCGGTGGGTGAGAGCGAGGCTGGCCATGCTCTGCT 98
Db 60598 GGCACCGCGTGGGCGGAGGCGCGCGAATGCTGCTGCT 60635

RESULT 15
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; BEST OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6
```

```
Query Match          52.1%; Score 52.6; DB 1; Length 28958;
Best Local Similarity 77.1%; Pred. No. 3.5e-06;
Matches 64; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
QY 19 GCTCCCGAGGAGCGGTGCAAGAGCTTCTGCGCCGACGCGCGTGGGGTGGAGCGAA 78
    ||||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db 15211 GCCCGGACGGTGTGCAAGGCCCTTCTGCGCGGAAGCCGACGCTGCGGCGGAA 15270
```

```
QY 79 GCGTGGCCATGCTCCTGCTCAA 101
    ||| || ||||||||| |
Db 15271 GGGCGGGGATGCTCTGCTCGA 15293
```

```
Search completed: November 6, 2002, 16:05:45
Job time : 206.625 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 : Search time 157 Seconds
(without alignments)
1448.738 Million cell updates/sec

Title: US-09-724-876-2_COPY_32070_32170

Perfect score: 101

Sequence: 1 agcgcgcctgcgagcctgcgc.....tgcgcacatctctctctcaaa 101

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	21	AA55887
2	101	100.0	71989	21	AA529349
3	58	57.4	3978	21	AA55785
4	58	57.4	12249	21	AA55840
5	58	57.4	18331	21	AA55857
6	57.8	57.2	20394	22	AA524892
7	57.6	57.0	672	22	AA589964
8	56.2	55.6	33529	23	AA517367
9	54.8	54.3	47981	22	AA530757

10	54.6	54.1	751	21	AA58914
11	54.4	53.9	15872	18	AA58715
12	54.4	53.9	15872	21	AA587283
13	54.4	53.9	43280	18	AA580413
14	54.4	53.9	65140	22	AA58186
15	54.4	53.9	125401	22	AA58186
16	53.2	52.7	16767	22	AA58339
17	53.2	52.7	50000	22	AA58833
18	53.2	52.7	50000	22	AA58836
19	53.2	52.7	80161	20	AA521501
20	52.6	52.1	28598	17	AA58769
21	52.6	52.1	28958	18	AA58956
22	52.6	52.1	28958	21	AA575299
23	52.6	52.1	49377	19	AA505287
24	52.6	52.1	53789	19	AA521187
25	52.6	52.1	671	22	AA58969
26	52.6	52.1	4306	22	AA590039
27	52.6	52.1	27541	22	AA58185
28	52.6	52.1	34071	22	AA590033
29	52.6	52.1	42717	22	AA590032
30	51.8	51.3	30690	21	AA592301
31	51.8	51.3	30690	22	AA592301
32	51.4	50.9	758	22	AA58927
33	51.4	50.9	1462	24	AA589970
34	51.4	50.9	4041	21	AA518435
35	51.4	50.9	5088	22	AA587300
36	51.4	50.9	11220	22	AA590038
37	51.4	50.9	36778	21	AA587298
38	51.4	50.9	37948	21	AA587318
39	51.4	50.9	38506	21	AA587285
40	51.4	50.9	38506	21	AA575633
41	51.4	50.9	11916	22	AA556001
42	51.2	50.7	12381	21	AA589279
43	50.2	49.7	44377	18	AA58381
44	50.2	49.7	44377	18	AA58508
45	50.2	49.5	29879	14	AA580414

ALIGNMENTS

RESULT 1	
AA55887	
ID	AA55887 standard; DNA: 68750 BP.
XX	
AC	AA55887:
XX	
DT	10-APR-2000 (first entry)
XX	
DE	Sorangium cellulosum 68.75 kb contig.
XX	
KW	Epithelone biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.
XX	
OS	Sorangium cellulosum.
XX	
FH	
FT	Location/Qualifiers
FT	1..1826
FT	/*tag= a
FT	/partial
FT	/product= "Partial Orf 1 protein (AA55880)"
FT	/note= "No initiation codon given in the specification"
FT	complement (1900..3171)
FT	/*tag= b
FT	/product= "Orf 2 protein (AA55851)"
FT	3415..5556
FT	/*tag= c
FT	/product= "Orf 3 protein (AA55852)"
FT	complement (5612..5992)
FT	/*tag= d
FT	/product= "Orf 4 protein (AA55853)"
FT	6226..6675
FT	/*tag= e
FT	CDS

DNA encoding ketos
Streptomyces venez
S. venezuelae vep
Tylactone synthase
Streptomyces nours
Streptomyces nours
S. spinosa DNA fra
S. spinosa DNA fra
S. spinosa DNA fra
DNA fragment of Sa
Sorangium cellulos
DNA sequence of So
The soraphen biosy
Amycolatopsis medl
Partial nucleotide
Nucleotide sequenc
Streptomyces nours
Nucleotide sequenc
Nucleotide sequenc
S. avermitilis ave
Streptomyces averm
Partial nucleotide
Contig 116 DNA enc
S. venezuelae macr
Nucleotide sequenc
S. venezuelae macr
S. venezuelae ptx
Nucleotide sequenc
Recombinant cosmids
Streptomyces averm
Streptomyces averm
Platenolide synthet
erya region of S.

```

FT      /product= "Orf 5 protein (AAV58584)"
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX      W09966028-A2.
XX      PD
XX      23-DEC-1999.
XX      PF
XX      16-JUN-1999; 99WO-EP04171.
XX      PR
XX      18-JUN-1998; 98US-0099504.
XX      PR
XX      24-SEP-1998; 98US-0101631.
XX      PR
XX      05-FEB-1999; 98US-0118906.
XX      PA
XX      (NOVS ) NOVARTIS AG.

```

```

PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX      WPI: 2000-097741/08.
XX      DR
XX      P-PSDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
XX      AAV58592, AAV58593, AAV58594.
XX      DR
XX      PT
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy.
XX      PS
XX      Claim 14; Page 87-104; 174pp; English.
XX      CC
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-Y58578) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX      CC
XX      SQ
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX      Query Match
XX      Best Local Similarity 100.0%; Score 101; Length 68750;
XX      Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      QY
XX      1 AGCGGCTGCGAGGCTGCTGCCGAGCGAGCGTTCGCGCCGACCGAC 60
XX      |||||||
XX      DB 37682 AGCGGCTGCGAGGCTGCTGCCGAGCGAGCGTTCGCGCCGACCGAC 3741
XX      QY
XX      61 GCGGTGGGGTGGAGCGAGCGTGGCCGATGCTGCTGCA 101
XX      |||||||
XX      DB 37742 GCGGTGGGGTGGAGCGAGCGTGGCCGATGCTGCTGCA 37782
XX      RESULT 2
XX      AAA29349
XX      ID AAA29349 standard; DNA; 71989 BP.
XX      AC
XX      AAA29349;
XX      DT
XX      12-SEP-2000 (first entry)
XX      DE
XX      Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX      KW
XX      Epoethione; polyketide synthase; epoa; epob; epoc; epod; epoe; epof;
XX      epog; epok; P450 oxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
XX      tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX      OS
XX      Sorangium cellulosum.
XX      FH
XX      Key Location/Qualifiers

```

```
FT CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /tag= c
FT /label= epocA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493
FT /tag= h
FT /label= epob_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137
FT /tag= x
FT /label= epocC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24059..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /tag= ap
```

```

FT      /note= "encodes ER5"
FT      30321..30869
FT      /*tag= aq
FT      /note= "encodes KR5"
FT      31077..31373
FT      /*tag= ar
FT      /note= "encodes ACP5"
FT      31440..32807
FT      /*tag= as
FT      /note= "encodes KS6"
FT      33018..34067
FT      /*tag= at
FT      /note= "encodes At6"
FT      34107..34676
FT      /*tag= au
FT      /note= "encodes DH6"
FT      35760..36641
FT      /*tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /*tag= aw
FT      /note= "encodes KR6"
FT      37470..37769
FT      /*tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /*tag= ay
FT      /label= "epoE gene"
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /*tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /*tag= ba
FT      /note= "encodes At7"
FT      41341..41922
FT      /*tag= Db
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCGGCTGCGAGGCTGCTCCCGAGCGAGGTGCAAGAGCTTTCGCGCGAGCCGAC 60
DB      32070 AGCCGGCTGCGAGGCTGCTCCCGAGCGAGGTGCAAGAGCTTTCGCGCGAGCCGAC 32129

QY      61 GGCCTGGGGTGGAGCGAAGGCTGCCATCTCTCTGCTCAA 101
DB      32130 GGCCTGGGGTGGAGCGAAGGCTGCCATCTCTCTGCTCAA 32170

RESULT 3
AAC55785
ID      AAC55785 standard; DNA: 3978 BP.
XX
AC      AAC55785;
XX
DT      19-JAN-2001 (first entry)
XX
DE      Type I polyketide synthase orf8.
XX
KW      Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
KM      anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KM      chronic obstructive pulmonary disease; respiratory inflammation;
KW      fungicide; pesticide; ds.
XX
OS      Streptomyces lavendulae.
XX
PN      MO200053737-A2.
XX
PD      14-SEP-2000.

```

```

XX      10-MAR-2000; 2000MO-0506394.
PF      12-MAR-1999; 99US-0266965.
XX
XX      (MINI) UNITY MINNESOTA.
PA      (SHER/) SHERMAN D H.
PA      (MAOY/) MAO Y.
PA      (VARO/) VAROGLU M.
PA      (HEMM/) HE M.
PA      (SHEL/) SHELTON P C.
XX
XX      Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
PI      WPI; 2000-601980/57.
XX
XX      Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT      cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT      the molecular basis of mitosome ring system biosynthesis
XX
XX      Example 1; Page 251-253; 399pp; English.
XX
XX      This invention relates to isolated and purified nucleic acid molecules
CC      from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC      natural products that contain a variety of functional groups, including
CC      amino benzocoumarone and axiridine ring systems. The S. lavendulae
CC      mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC      spanning 55kb of DNA. The invention includes an expression cassette
CC      comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC      and host cells transformed with the cassette. The nucleotide, and protein
CC      sequences and the transformed host cells of the invention result in
CC      antitumor, anti-inflammatory, cytostatic, immunomodulatory, and
CC      antibiotic activities. The nucleotide sequences are used to elucidate the
CC      molecular basis for the biosynthesis of the mitosome ring system, as well
CC      as to engineer the biosynthesis of novel natural products, e.g.,
CC      antibiotics, anti-inflammatory agents, anti-cancer agents,
CC      immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC      obstructive pulmonary disease as well as other disease involving
CC      respiratory inflammation, or cholesterol-lowering agents or as crop
CC      protection agents (e.g., fungicides or insecticides) as well as
CC      biopolymers, e.g., in packaging or biomedical applications, or to engineer
CC      PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC      AAB34483-B32342 represent mitomycin biosynthetic gene cluster DNA
CC      sequences and encoded proteins. Sequences AAC55812-C55814,
CC      AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC      cloning of the mitomycin biosynthetic genes.
XX
SQ      Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 other;

Query Match      57.4%; Score 58; DB 21; Length 3978;
Best Local Similarity 74.5%; Pred. No. 6.1e-07;
Matches 73; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY      1 AGCCGGCTGCGAGGCTGCTCCCGAGCGAGGTGCAAGAGCTTTCGCGCGAGCCGAC 60
DB      742 AGCCGGCTGCGAGGCTGCTCCCGAGCGAGGTGCAAGAGCTTTCGCGCGAGCCGAC 801

QY      61 GGCCTGGGGTGGAGCGAAGGCTGCCATCTCTCTGCT 98
DB      802 GGCACCACTGTGTCGAGGCGCGGCGCTGCTGCT 839

RESULT 4
AAC55840
ID      AAC55840 standard; DNA: 12249 BP.
XX
AC      AAC55840;
XX
DT      19-JAN-2001 (first entry)
XX
DE      Complete Mitomycin ORF 1-9 nucleotide sequence.
XX
KW      Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;

```

[illegible]

ID	AA055857	standard; DNA; 18331 BP.
AC	AA055857	
DT	19-JAN-2001	(first entry)
XX		
DE	Complete nucleotide sequence of the mitomycin gene cluster.	
XX		
XX	Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;	
XX	anti-inflammatory; immune-enhancer; immunosuppressant; asthma;	
XX	chronic obstructive pulmonary disease; respiratory inflammation;	
XX	fungicide; pesticide; ds.	
OS	Streptomyces lavendulae.	
PN	WO200053737-A2.	
XX		
PD	14-SEP-2000.	
XX		
PF	10-MAR-2000; 2000WO-US06394.	
XX		
PR	12-MAR-1999; 99US-0266965.	
XX		
PA	(MINU) UNIV MINNESOTA.	
PA	(SHER/) SHERMAN D H.	
PA	(MAOY/) MAO Y.	
PA	(VARO/) VAROGU M.	
PA	(HEMU/) HE M.	
PA	(SHEL/) SHELDON P C.	
XX		
PI	Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;	
XX		
DR	WPI; 2000-601980/57.	
XX		
PT	Novel nucleic acid molecule comprising mitomycin biosynthetic gene	
PT	cluster useful for cloning mitomycin biosynthetic genes for elucidating	
XX	the molecular basis of mitosome ring system biosynthesis	
XX		
PS	Disclosure; Figure 21; 39pp; English.	
XX		
CC	This invention relates to isolated and purified nucleic acid molecules	
CC	from the mitomycin biosynthetic gene cluster. Mitomycins are a group of	
CC	natural products that contain a variety of functional groups, including	
CC	amino benzoquinone and axiridine ring systems. The S. lavendulae	
CC	mitomycin biosynthetic gene cluster comprises 47 mitomycin genes	
CC	spanning 55kb of DNA. The invention includes an expression cassette	
CC	comprising a mitomycin biosynthetic gene operably linked to a promoter,	
CC	and host cells transformed with the cassette. The nucleotide, and protein	
CC	sequences and the transformed host cells of the invention result in	
CC	antitasthmatic, antiinflammatory, cytostatic, immunomodulatory, and	
CC	antibiotic activities. The nucleotide sequences are used to elucidate the	
CC	molecular basis for the biosynthesis of the mitosome ring system, as well	
CC	as to engineer the biosynthesis of novel natural products, e.g.	
CC	antibiotics, anti-inflammatory agents, anti-cancer agents,	
CC	immune-enhancers, immunosuppressants, agents to treat asthma, chronic	
CC	obstructive pulmonary disease as well as other disease involving	
CC	respiratory inflammation, or cholesterol-lowering agents or as crop	
CC	protection agents (e.g. fungicides or insecticides) as well as	
CC	biopolymers, e.g., in packaging or biomedical applications, or to engineer	
CC	PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and	
CC	AA832485-332542 represent mitomycin biosynthetic gene cluster DNA	
CC	sequences and encoded proteins. Sequences AAC55812-C55814,	
CC	AAC55850-C55865 and AAC55862-C55869 represent PCR primers used in the	
CC	cloning of the mitomycin biosynthetic genes.	
XX		
SO	Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;	
XX		
Query Match	57.4%;	Score 58; DB 21; Length 18331;
Best local similarity	74.5%;	Pred. No. 6.5e-07;
Matches 73;	Conservative 0;	Mismatches 25; Indels 0; Gaps 0;
1	AGCCGCGTCGAGGCGCTGGCTCCGACGACGCGTGCACAGAGCTTCGCGCCGACCGAC 60	


```

XX 12-MAR-2002 (first entry)
XX
XX DNA sequence of S. cellulosum polyketide synthase cosmid, PKOS28-26.
DE
XX
XX Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
XX acyl transferase domain; dehydratase domain; ketoreductase domain;
XX acyl carrier protein domain; PKOS28-26; ds.
XX
XX Sorangium cellulosum.
XX
XX US6280999-B1.
XX
XX 28-AUG-2001.
XX
XX 31-AUG-1998; 98US-0144085.
XX
XX 22-JAN-1998; 98US-0010809.
XX
XX (KOSA-) KOSAN BIOSCIENCE.
XX
XX Gustafsson C, Belach MC, Ashley G, Julien B, Ziemann R;
XX
XX WPI; 2001-606536/69.
XX
XX
XX Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX polyketide open reading frame encoding modules with one or more domains
XX such as ketosynthase, acyl transferase and acyl carrier protein domains
XX
XX
XX Claim 4; Fig 1; 72pp; English.
XX
XX
XX The present invention relates to the isolation of novel Sorangium
XX cellulosum polyketide synthases (PKS), and the polynucleotide sequences
XX encoding them. The polyketide synthases include catalytic domains such
XX as ketosynthase domain, acyl transferase domain, dehydratase domain,
XX ketoreductase domain and acyl carrier protein domain. A host cell
XX comprising a PKS ORF (open reading frame) which encodes one or more
XX more PKS domains is useful for producing polyketide synthases from which
XX polyketides can be produced. The host cells are useful for constructing
XX a library, where each individual colony of the library represents a
XX colony with the ability to produce a particular PKS synthase and
XX ultimately a particular polyketide. The polyketides produced by these
XX colonies can be used collectively in a panel to represent a library or
XX may be assessed individually for activity. Colonies in the library are
XX also induced to produce the relevant synthases and thus to produce the
XX relevant polyketides to obtain a library of candidate polyketides which
XX can be screened for binding to desired targets such as receptors,
XX signalling proteins, etc. The present sequence represents the DNA
XX sequence of cosmid PKOS28-26 which encodes one or more domains of
XX S. cellulosum PKS.
XX
XX Note: The present sequence is said to encode the functional domains
XX of S. cellulosum PKS which correspond to domains or domain subsets of
XX the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
XX Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other:
XX
XX Query Match 55.68; Score 56.2; DB 23; Length 33529;
XX Best Local Similarity 72.38; Pred. No. 2.1e-06;
XX Matches 73; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
XX
XX
XX 1 AGCCGCTGCGAGCTGCTCCGAGGAGCGTGAAGACCTCTGCGCCGACGCGAGC 60
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 376 AGCCGCTCAAGGGGATGCGCCGCGAGCGCCGCTGCAAGACCTCTCTGCGCGAGCTGAC 435
XX
XX 61 GGCCTGGGGTGAGCAGAGGCTGCGCATGCTCTCTGCTCAA 101
XX | | | | | | | | | | | | | | | | | | | | | |
XX Db 436 GGCCTCAACCTGTCTCGAGGGGTGCGGATGCTGGTGA 476
XX
XX
XX RESULT 9
XX AAF30757
XX ID AAF30757 standard; DNA; 47981 BP.
XX
XX
XX AAF30757;
XX
XX 21-JUN-2001 (first entry)
XX
XX Micromonospora megalomicin biosynthetic gene cluster.
DE
XX
XX Megalomicin; meg gene; polyketide synthase; antibiotic;
XX methylide; antiparasitic; ds.
XX
XX Micromonospora megalomicin subsp. nigra.
XX
XX Key
XX CDS
XX
XX Location/Qualifiers
XX complement (1..144)
XX /tag= a
XX /partial
XX /gene= "megBVI(megI)"
XX /product= "TDP-4-keto-6-deoxyglucose-
XX 2,3-dehydratase"
XX /note= "encodes AAB82201"
XX 928..2061
XX /tag= b
XX /gene= "megDVI"
XX /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
XX /note= "encodes AAB82202"
XX 2072..3382
XX /tag= c
XX /gene= "megDI"
XX /product= "TDP-megosamine glycosyltransferase"
XX /note= "eryCIII homologue; encodes AAB802203"
XX 3462..4634
XX /tag= d
XX /gene= "megI"
XX /product= "mycarose O-acyltransferase"
XX /note= "encodes AAB82204"
XX 4651..5775
XX /tag= e
XX /gene= "megDII"
XX /product= "TDP-3-keto-6-deoxyhexose
XX 3-aminotransaminase"
XX /note= "eryCI, DnrI homologue; encodes AAB82205"
XX 5822..6595
XX /tag= f
XX /gene= "megDIII"
XX /product= "daunosaminyl-N,N-dimethyltransferase"
XX /note= "eryCVI homologue; encodes AAB82206"
XX 6592..7197
XX /tag= g
XX /gene= "megDIV"
XX /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
XX /note= "eryBVIII, dnmU homologue; encodes AAB82207"
XX 7220..8206
XX /tag= h
XX /gene= "megDV"
XX /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
XX /note= "eryVIV, dnmV homologue; encodes AAB82208"
XX complement (8228..9220)
XX /tag= i
XX /gene= "megBII-1(megDVII)"
XX /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
XX /note= "encodes AAB82209"
XX complement (9226..10479)
XX /tag= j
XX /gene= "megBV"
XX /product= "TDP-mycarose glycosyltransferase"
XX /note= "encodes AAB82210"
XX complement (10483..11424)
XX /tag= k
XX /gene= "megBIV"
XX /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
XX /note= "encodes AAB82211"
XX 12181..22821
XX /gene= "mega"

```


DE	Streptomyces venezuelae polyketide synthase vep ORF1.
XX	
KW	polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW	polyhydroxybutyrate; biodegradable polymer; vep gene;
KW	metabolic engineering; ss.
XX	
OS	Streptomyces venezuelae.
XX	
FT	Key
FT	Location/Qualifiers
FT	CDS
FT	20..13912
FT	/*tag= a
FT	14056..14136
FT	/*tag= b
FT	14148..15827
FT	/*tag= c
XX	
XX	WO9722711-A1.
XX	
XX	26-JUN-1997.
XX	
XX	18-DEC-1996; 96WO-US20119.
XX	
XX	19-DEC-1995; 95US-0008847.
XX	
XX	(MINU) UNIV MINNESOTA.
XX	
PI	Sherman DH, Williams MD, Xue Y;
XX	
DR	WPI; 1997-341701/31.
DR	P-PSDB; AAMI9629-30 AND AAMW0918.
XX	
PT	Expression cassettes for production of polyhydroxyalkanoate(s) -
PT	provide wide range of biodegradable polymers for medical or
PT	industrial use
XX	
XX	Claim 54; Fig 23; 91pp; English.
XX	
PS	Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
XX	
CC	synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC	also AAMI9629-30 and AAMW0918). It contains 5 PKS modules, with a 5'
CC	loading module and a 3' end domain. Each of the sequenced modules
CC	includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC	reductase and an acyl carrier protein domain. The gene cluster was
CC	cloned using a heterologous hybridisation strategy from a genomic
CC	DNA library. A novel expression cassette encoding the first module
CC	from the vep gene cluster and module 7 from the Streptomyces ty1P
CC	gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC	activity and can be used for PHA prodn. in host (esp. insect) cells
CC	for use as a biodegradable polymer.
XX	
SO	Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;
	Query Match 53.9%; Score 54.4; DB 18; Length 15872;
	Best Local Similarity 72.9%; Pred. No. 6.2e-06;
	Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY	3 CGCGCTGCAGAGCCCTGGCTCCCGAGAGCAGTGCAGAGCTTCGCGCCAGCGAGCG 62
DB	9136 CCGGCAAGAAAGGGGCTGGCCGCGAGCGCGCTGCAGAGCGTTCTCGCGCCGACGG 9195
QY	63 CGTGGAGGAGAGCGAAGGCTGGCCATCTCTCTGCT 98
DB	9196 CACCGCTGGGCGCGAGGCGCTGGCTGCTCTCTCT 9231
	RESULT 12
ID	AAZ87283
	AAZ87283 standard; DNA; 15872 BP.
XX	
AC	AAZ87283;
XX	
DT	05-JUN-2000 (first entry)
XX	

DE S. venezuelae vep ORF 1, SEQ ID NO:1.
 XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanote monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent; ds.
 XX Streptomycetes venezuelae ATCC15439.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 20..13912
 FT /tag= a
 FT /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"
 FT CDS 14056..14151
 FT /tag= b
 FT /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"
 FT CDS 14167..15827
 FT /tag= c
 FT /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"
 FT
 PN MO200000620-A2.
 XX
 PD 06-JAN-2000.
 XX
 XX 25-JUN-1999; 99WO-US14398.
 PF
 XX 26-JUN-1998; 98US-0105537.
 PR
 XX (MINT) UNIT MINNESOTA.
 PA
 PI Sherman DH, Liu H, Xue Y, Zhao L;
 DR WPI: 2000-160679/14.
 P-PSDB: AAV77177, AAV77178, AAV77199.
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 PT synthesis of methymycin and pikromycin -
 XX
 PS Example 3; Figure 23; 438bp; English.
 XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the ergC gene cluster of *Saccharopolyspora erythraea* or
 CC Streptomycetes antibiotics. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanote (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants. The
 CC present sequence represents a Streptomycetes venezuelae ATCC 15439 DNA
 CC sequence, designated vep ORF 1 in the specification, which actually
 CC contains 3 open reading frames, which encode proteins AAV77177-V77178 and
 CC AAV77199. The vep ORF 1 protein is defined in the specification as a PHA
 CC monomer synthase.
 XX
 XX Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

Query Match 53.9%; Score 54.4; DB 21; Length 15872;
 Best Local Similarity 72.9%; Pred. No. 6, 2e-06;
 Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 Oy 3 CCGGCTGCAGAGCCCTGGCTCCACAGGAGCGTCCAGAGCTTTCGGCCGACCGACGG 62
 ||||| | | ||||| ||||| | ||||| ||||| ||||| ||||| ||||| |||||
 Db 9136 CCGGAGAGAGGGGCGTCCGCCGACGCGCTCCAGAGCGTTCGCGCCGCGACCG 9195
 ||||| | | ||||| ||||| | ||||| ||||| ||||| ||||| ||||| |||||
 Oy 63 CGTGGGCTGGAGCGGAGCGCTGCCCATGCTCTGCT 98
 | || | | || | | | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 9196 CACCGGCTGGGCGGAGGCGCTGCTGCTCT 9231
 ||||| | | ||||| ||||| | ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 13
 AAT80413
 ID AAT80413 standard; DNA; 43280 BP.
 XX
 AC AAT80413;
 XX
 XX 27-FEB-1998 (first entry)
 DT
 XX
 DE Tylactone synthase gene cluster.
 KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
 KW polyketide; tylactone synthesis; antibiotic; tylosin; ss.
 XX
 OS Streptomycetes fradiae.
 XX
 PH Key Location/Qualifiers
 FT CDS 816..14243
 FT /tag= a
 FT /transl_except= (pos: 816..818, aa: Met)
 FT /note= "ORF1 encodes protein shown in AAW22601"
 FT CDS 14351..19945
 FT /tag= b
 FT /transl_except= (pos: 14351..14353, aa: Met)
 FT /note= "ORF2 encodes protein shown in AAW22602"
 FT CDS 20010..31199
 FT /tag= c
 FT /transl_except= (pos: 20010..20012, aa: Met)
 FT /note= "ORF3 encodes protein shown in AAW22603"
 FT CDS 31232..36067
 FT /tag= d
 FT /note= "ORF4 encodes protein shown in AAW22604"
 FT CDS 36249..41774
 FT /tag= e
 FT /note= "ORF5 encodes protein shown in AAW22605"
 FT
 XX EP791655-A2.
 PN
 XX 27-AUG-1997.
 DR
 PF 19-FEB-1997; 97EP-0301056.
 XX
 XX 22-FEB-1996; 96US-0012078.
 PR
 XX (ELIL) LILLY & CO ELI.
 PA
 PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
 DR WPI: 1997-418046/39.
 P-PSDB: AAW22601-W22605.
 XX
 PS Claim 2; Pages 8-66; 220bp; English.
 XX This sequence represents the tylactone synthase gene cluster of the
 CC invention. This sequence is also referred to as the tylG gene, and was
 CC isolated from Streptomycetes fradiae. This sequence encodes multifunctional
 CC proteins which direct the synthesis of the polyketide tylactone, isolated

CC from Streptomyces fradiae. Tylactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.

XX Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

Query Match 53.9%; Score 54.4; DB 18; Length 43280;

Best Local Similarity 72.9%; Pred. No. 6.4e-06; Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CCGGCTGCGAGGCGCTGCTCCGACGACGAGCTTCTGCGCGACCGACGG 62

Db 36998 CCGGCAAGAAGGCGCTGCGCGACGCGCGTCTGCGCGCGCGACGG 37057

QY 63 CGTGGGTGAGACGAGCGCTGCGCATGCTCTGCT 98

Db 37058 CACGCGCTGGGCGCGAGCGCTGCGCTGCTCT 37093

RESULT 14

AAD17184 standard; DNA; 65140 BP.

AC AAD17184;

DE 29-NOV-2001 (first entry)

XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

KM antifungal; antibiotic; nys1; ds.

XX Streptomyces noursei.

XX Location/Qualifiers
FH complement (1..1035)

FT /tag= a
FT /product= "NysD2 partial protein"

FT /note= "CDS does not include stop codon"

FT /tag= b
FT /product= "NysD1 protein"

FT /tag= c
FT /product= "NysA protein"

FT /tag= d
FT /product= "NysB protein"

FT /tag= e
FT /product= "NysC protein"

FT /tag= f
FT /product= "NysE protein"

FT /tag= g
FT /product= "NysR1 protein"

FT /tag= h
FT /product= "NysR2 protein"

FT /note= "CDS does not include start codon"

FT /tag= i
FT /product= "NysR3 protein"

FT /tag= j
FT /product= "NysR4 (short) protein"

FT /note= "CDS does not include start codon"

FT /tag= k
FT /product= "NysR5 protein"

FT /note= "CDS does not include start codon"

FT complement (62551..63615)

FT /tag= l
FT /product= "ORF2 protein"

FT /note= "CDS does not include start codon"

FT /tag= m
FT /product= "ORF1 protein"

XX WO200159126-A2.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB00509.

XX 08-FEB-2000; 2000GB-0002840.

XX 10-APR-2000; 2000GB-0008786.

XX 14-APR-2000; 2000GB-0009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

XX (SMTF) SINTER STIPELSEN IND TEK FORSK.

XX (ALPH-) ALPHARMA AS.

XX (SINV-) SINVENT AS.

XX (DZIE-) DIEGLEWSKA H.

XX (ZOTC/) ZOTCHEV S B.

XX (SEKU/) SEKUROVA O N.

XX (FJAE/) FJAEVRIK E.

XX (BRAU/) BRAUTASET T.

XX (STRO/) STROM A R.

XX Zorchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI: 2001-557614/62.

XX P-PDB: AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,

XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10137.

XX New nystatin polyketide synthase polynucleotides and polypeptides,

XX useful as antibiotics and antifungals -

XX Claim 2; Page 116-151; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

XX cluster encoding a modular type I polyketide synthase (PKS) enzyme

XX involved in the biosynthesis of the macrolide antibiotic nystatin.

XX The nystatin PKS is useful as antifungal antibiotics. The present

XX sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

Query Match 53.9%; Score 54.4; DB 22; Length 65140;

Best Local Similarity 72.9%; Pred. No. 6.5e-06; Matches 70; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CCGGCTGCGAGGCGCTGCTCCGACGACGAGCTTCTGCGCGACCGACGG 62

Db 34048 CCGGCAAGAAGGCGCTGCGCGACGCGCGTCTGCGCGCGCGACGG 34107

QY 63 CGTGGGTGAGACGAGCGCTGCGCATGCTCTGCT 98

Db 34108 CGTGGGTGAGACGAGCGCGTCTGCGCATGCTCT 34143

RESULT 15

AAD17186 standard; DNA; 125401 BP.

AC AAD17186;

DE 29-NOV-2001 (first entry)

XX Streptomyces noursei nystatin PKS gene cluster DNA.

XX	OS	Streptomyces noursei.	
XX	XX		
XX	XX	Location/Qualifiers	
XX	XX	6337..34771	
XX	XX	/tag- a	
XX	XX	/product- "NysI complete protein"	
XX	XX	34792..51099	
XX	XX	/tag- b	
XX	XX	/product- "NysJ protein"	
XX	XX	51155..57355	
XX	XX	/tag- c	
XX	XX	/product- "NysK protein"	
XX	XX	57503..58687	
XX	XX	/tag- d	
XX	XX	/product- "NysL protein"	
XX	XX	complement (58786..58980)	
XX	XX	/tag- e	
XX	XX	/product- "NysM protein"	
XX	XX	/note- "CDS does not include start codon"	
XX	XX	complement (60238..61296)	
XX	XX	/tag- g	
XX	XX	/product- "NysD2 complete protein"	
XX	XX	120628..121308	
XX	XX	/tag- h	
XX	XX	/product- "NysR4 (long) protein"	
XX	XX		
XX	XX	WO200159126-A2.	
XX	XX		
XX	XX	16-AUG-2001.	
XX	XX		
XX	XX	08-FEB-2001; 2001WO-GB00509.	
XX	XX		
XX	XX	08-FEB-2000; 2000GB-0002840.	
XX	XX	10-APR-2000; 2000GB-0008786.	
XX	XX	14-APR-2000; 2000GB-0009387.	
XX	XX		
XX	XX	(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIG.	
XX	XX	(SMTF) SINTEF STIPELSEN IND TEK FORSK.	
XX	XX	(ALPH-) ALPHARMA AS.	
XX	XX	(SINV-) SINVENT AS.	
XX	XX	(DZIE/) DZIELENSKA H.	
XX	XX	(ZOTC/) ZOTCHEV S B.	
XX	XX	(SEKU/) SEKUROVA O N.	
XX	XX	(FJAE/) FJAEERVIK E.	
XX	XX	(BRAU/) BRAUTASET T.	
XX	XX	(STROM/) STROM A R.	
XX	XX		
XX	XX	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;	
XX	XX	Valla S, Ellingsen TE, Sletta H, Gulliksen O;	
XX	XX	WPI: 2001-557614/62.	
XX	XX	P-RSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,	
XX	XX	AAE10149, AAE10150.	
XX	XX		
XX	XX	New nystatin polyketide synthase polynucleotides and polypeptides,	
XX	XX	useful as antibiotics and antifungals -	
XX	XX	Claim 1; Page 188-254; 266pp; English.	
XX	XX		
XX	XX	The present invention relates to the cloning and sequencing of the gene	
XX	XX	cluster encoding a modular type I polyketide synthase (PKS) enzyme	
XX	XX	involved in the biosynthesis of the macrolide antibiotic nystatin.	
XX	XX	The nystatin PKS is useful as antifungal antibiotics. The present	
XX	XX	sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.	
XX	XX		
XX	XX	Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;	

Query Match	Similarity	Score	5d.4:	DB	22:	length	125401:
Best Local	Similarity	72.9%	Pred.	No. 6.7e-06:			
Matches	70:	Conservative	0:	Mismatches	26:	Indels	0:
QY	3	CCGGCTGGCAGGCGCTGGGTC	CCCGATGGGAGAGGAGCAAGACTTCTCGGCGCAGCCGACGG	62			
Db	94309	CCGGCAGGCGCGGACTGGGCCCCCGAGGCGCGCTCCAGAGCCGTTCTCCGATCCGCGCCGACCG	94368				
QY	63	CGTGGGCTGAGACGCAAGGCTGGCCATGCTCTCTGCT	98				
Db	94369	CGTGGCTGTCGTCGAGGCGGCTCGGAATGCTGCTCT	94404				

```
Search completed: November 6, 2002, 12:18:14
Job time : 255 secs
```


AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B.

TITLE Cloning and heterologous expression of the epothillone gene cluster

JOURNAL Science 287 (5453), 640-642 (2000)

MEDLINE 20115953

PUBLISHED 10649995

REFERENCE 2 (bases 1 to 58733)

AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.

TITLE Isolation and characterization of the epothillone biosynthetic gene cluster from *Sorangium cellulosum*

JOURNAL Gene 249 (1-2), 153-160 (2000)

MEDLINE 20293058

PUBLISHED 10831849

REFERENCE 3 (bases 1 to 58733)

AUTHORS Julien, B.

TITLE Submitted

JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES

source Location/Qualifiers

CDS

/organism="Polysangium cellulosum"

/strain="SMP4"

/db_xref="taxon:56"

<1..992

/note="OrfA"

/codon_start=3

/transl_table=11

/product="putative transposase"

/protein_id="AA062878.1"

/db_xref="GI:7453555"

/translation="VRGHEAFADFGVPRVLLYDNLKNAVERHGDAIRFHPILLAL
SADYFEPRVAVARGNEKGRVERAIRYREGFEPARVADIGDNLROATEMTSSAL
DRSNVEDARTVRAQFDDERSVLLRHDPDPDHEVEVEVEKTYARPDENDYSAL
DTRRTLVVADLSQVRIADGNQIVATHRSDRGQOIEPEHLORLDEKRAEHR
GIDRLARRASQAFRLIYERGDVNGSIAIRLLOLVANGAELEAEVLEEDTI
HIGAVROYIDRRSRERHPPPISTVYNGEHALVVTTHSTTYTALKKDP"

989..1501

/note="OrfB"

/codon_start=1

/transl_table=11

/product="putative transposase"

/protein_id="AA062879.1"

/db_xref="GI:7453556"

/translation="MTDLPTETDKRLSLGFLGLACMEQLADKPMLEEVLAIEERE
BHKRLERLKNKRYAAAFKPMTPDSDSMPKIDREAVDLDYSDRAIDLFEVYTRRD
AOKPLLSLNKAFADMGQVFPAAACVVTLVDRLVRAEYIEIASYRLKKEKELNAT
RTKRRTRKH"

CDS

1998..6263

/gene="epoc"

1998..6263

/gene="epocA"

/note="type I polyketide synthase; epothillone PKS loading module"

/codon_start=1

/transl_table=11

/product="EpoA"

/protein_id="AA062880.1"

/db_xref="GI:7453557"

/translation="MADRIERAADPTAIYAGCRLRGVIDLSGFMTLLGSDTY
GOVPERMDAAMPDDLDAPKTYVTRASFSLSDVACDASFEFGISIPREALMDPAHR
LLEECWELNMAAIPASLVGTETGVFIGIPSEYEAALPRATASAEIDALMDGATG
PSVAGRIISYVGLRGKCAVADTAVSSSLVAHGLACOSLRSECSALAGVSLMSP
SLVWLSKTRALATDGRCKAFSADGFGRGCAVVVLRSLSGARADGRLIAYIRG
SAINDGASGGLTVNGSSEITVLRALADAGCAASVYGAHGTCTTIDPTICQA
LNAYIGLRDVAATPLLIGSVKTNLGHPEYASGITLKVYVLSLQGGIIPAHHAADLN
PRISWGLRLVTRARTPWPMTNTPRAGVSSFGSGTNAHVLEAPACTPPAPE
PRAELLVISARTAAALDAHAARLDHLETTYSCIGDAVSLATTSAMERHLAAVAT
SSEGRALDAAGAGOTPGVVRGJADSRGCLAFITGOGAQTLLGMGRGLYDVPAF
REAPDVRIRFNOELDRPLIREYMAEPASVDAALLDQTAFTQPALFTFEVYALMRS
WCEPELVAGHSIGELVACVAGSLEDVATVACARIMALPAGAMSTIAPEA
DVAAAVARHAASVSTIAVNGPDQVIAAGQGVPIIAAMAMARGAKTRALHVSFHS
PLAAMPLEAFGRVAVESVYRPSIVLVSNLSGKAGTDEVSPGVWRARAREVVRADG
VKAHHAAGATFEVEGPKSTLGLVPACLIPARPALILASSRAGRDEPATVLEALGLM

gene

CDS

gene

CDS

AVGLSVNAGLEPPSGRRVPLPTYPWQBERYVIDTKADAARAGDRAPGACHDEVEKG
CAVRGCDGRRSALDRPPESGRKEVREARPELEIDEDEVDRLRTYERAPG
LGEVEIADNAGLSFNDVOLAGVDDLPKGNPPLILGEGCGRIYAVEGVNAGV
VGOPIVATISAGAFTHHTTSAVILYPRQALISATPAAAPAVITAWALDGIARLOP
GERVLIHATGCGVGLAAYQMAOHVCAEPHATACGTEKARAYILESLGVRVYSQSDREY
ADVRAWTGEGEVVNLNSLGEIDKSNLRSRFRVELKROCIADANQISGRDFR
NLSFSLVDRGMMLERPARVRALEFELGILAAAGVTPPATITLPIAVADAFRSMQ
AQHLKLVLTLDDEVOAIRPIPHAGAGSTGDRDLDRDLASAAPARAALAEALRTQ
VSOVLRPEIKVAGALFTRLGDSIMAVEIRNRITAEASTKLKSLTFLSTSPNIALTL
ONLDAATPALSLEVAENLRAGVQDSFVSSGADQDEIITAL"

6260..10492

/gene="epocB"

6260..10492

/gene="epocB"

/note="non ribosomal peptide synthetase; epothillone PKS module 1"

/codon_start=1

/transl_table=11

/product="EpoB"

/protein_id="AA062881.1"

/db_xref="GI:7453558"

/translation="MTINQNLNLEHOGVTKLAQDSERLQIQAPKNALPNILARISEH
KSTIITMLRQRLPASITYPAPAREHVRPPLIDIGSYLGRGAFVYSGIHAREYD
CTDLVVARLSRAFRKVARHMDLRHATLPDMAQVTEKVDADIEITDLRGDRSTREA
RLVSLRDAMSHRIYDTERPPYHVAVVARDQQLRVLSIDLINVDGLSLIIFKDWL
SEYEDPETSLPYLESLDYRDIYLALESRRKSAHORSMQYRKARVAELPPPMKAD
PSTLREIRFRHTEQMLPSDSWSRLKORGEGLPTGYILAAFSVIGRMSAPFETL
NITLFNRLVPHRVNDITGDTSMVLLIDITTRDKSPFQRAKRIQOELWEMMDHVS
CIENOREARVILGIRGALFPVYLISALNQQOVYTSQIRGCTGTPVYSTQIPOLLIDH
OLYEHGDVILAMDIVDGFPPDLDDMLAEVATLRLTLEPMSQDRCLPQAOLE
AASANENLNLSEHTLHGLFAARVQPLPMOLAVVASAKTLYLELSRSRLAEKL
BOGARNPLVAVNAGMEQVAVLAIVLESQAAYPIDADIAPAEIHLHLHGEVLY
LTQPMIDRACKSVMPGIGLIVLSDAGVEDGQGLPMPVLOTSDLAIVYTGSGTGLPK
GVMDIRGKAVNMTILDINRFELPGEDRVLSLEFGLSDVYDVGILAAAGTIVPDA
SKLRDPAHMAALIEEKYTVNNSVAPALMBMVEHSEGGDSIASITLSLGMQPIY
GUPGLQAIRREVSYTISGAGTASNTSITCPVRNVDPSMASIPYGRPLRQGTAVLD
EALERPVAWPGQLYIGGVGLALGYMRQDEETRNSFLVHPBTGERLYTGDGLRPLD
GNIEFMRGENDQIKLGRYVELGEIETELKSHIPNRADVIVPVGDANKLILAYVP
EGTRRRAEODASLKTERRDARAHAAKADSGDRVQFKLARHGLRDLGKFPVLD
TGLVPRAGLDVYARRSRVTFLEAPITPFVEGRFLSCGLSSVEPGGALPFRPVSAG
STYPQVQYAAKSGRIEGVDGCFYHYHFEHRLIKVSHGIERGAHVQNDVDDEA
RGLFVGRHIDALIESLYSGLSRRECLLEGYQAQLMEDAPSCNIGVCVQPDPOYR
PVLDIRHSDVYVHGMIGRVDPROFQCTLQODSSPRATRTGAPPGDHQFADILRD
FLKRTLPEYMPVYVEVLDALPLTNSNGVDRKALREKRTDSSPRSHHTAPRDLEEI
IVAAVREVLGLEVVGLQOSFVLDGATSLHIYMRSLQKRLDREIATITELTQYPNLGS
LASGLRDSKDLDPNMQDRVEARKGRRS"

gene

CDS

/note="type I polyketide synthase; epothillone PKS module 2"

Query Match	Best Local Similarity	100.0%	Score 101;	DB 1;	Length 58733;
Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1	AGCGGGCTGCAGAGGCGCTGCTCCGAGCGAGGCGAGGTCCTCTCGCGCGAGCGGAC	60			
Db 32070	AGCGGGCTGCAGAGGCGCTGCTCCGAGCGAGGCGAGGTCCTCTCGCGCGAGCGGAC	32129			
QY 61	GCGGTGGGGTGAGAGCGAGGCTGCGCCATGCTCTGCTCA 101				
Db 32130	GCGGTGGGGTGAGAGCGAGGCTGCGCCATGCTCTGCTCA 32170				
RESULT 2					
AF210843					
LOCUS					
DEFINITION	Sorangium cellulosum strain So c690 epothilone biosynthesis gene				
ACCESSION	AF210843				
VERSION	AF210843.1				
KEYWORDS					
SOURCE					
ORGANISM	polyangium cellulosum. Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.				
REFERENCE					
AUTHORS	1 (bases 1 to 68750) Molnar,I., Schupp,T., Ono,M., Zipfke,R., Mliamow,M., Nowak-Thompson,B., Engel,N., Toupet,C., Strahmann,A., Cyr,D.D., Goriach,J., Mayo,J.M., Hu,A., Golf,S., Schmid,J. and Ligon,J.M. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So c690				
TITLE					

JOURNAL	Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE	20130945
PUBMED	10662695
REFERENCE	2 (bases 1 to 68750)
AUTHORS	Moinar, I.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box 12257, Research Triangle Park, NC 27709, USA
FEATURES	Location/Qualifiers
source	1..68750 /organism="Polyangium cellulosum" /strain="So ce90" /db_xref="taxon:56" <1..1826 /note="ORF1" /codon_start=3 /transl_table=1 /product="unknown" /protein_id="AAF26904.1" /db_xref="GI:6724238" /translation="ASDIALFARATSAVLDDGHGRATERHVLAEAGIEDLALREH LRIOEGSPHCWICIDGLTVLEILADHDPDPLASISPHNRSLRHPMTSDALVGPALV RMILAGARGBRIRYREERERARACQARLILAAAPCPALPREDDANCPALPV MSPEVAERLRIRASTYTPPELACALALMIGTGCSPSGIPATEMLEPENLLTGELTP AIIAASAPGTSEEARLKGARLFASWEVSSKSSQLGNI PALMERLPT IVRANGNADN LSREREAIAIAEEVRLRAOPAPPAAGAGLAVGVSSSGLSGLVTDGALYSGDND IVMPGPRISPVVLLAGTDPEFFELAPLSOMLFYAHNAGTI SKVLEGSPLIYMAN OARMSIVHARGFAMVANOAMPDPERGAPFYVORSI IMEFHEPTPLCHPEPASAS LACDEHLNCELSAGRLLELRHHHRGAPSRATYGERPIATATYPSITLANTHYL WADDRKALIGVDKRTGVEPLVLAETHPPAHVSEDRDIFALIGDPSKDMVHEIR SGASTVADYQORDMDRPPDVMILNRNGLEFTTNRLITLAS"
CDS	complement(1900..3171) /note="ORF2: similar to hypothetical proteins from Mycobacterium tuberculosis and streptomycetes coelicolor and to D2-peptidases" /codon_start=1 /transl_table=1 /product="unknown" /protein_id="AAF26905.1" /db_xref="GI:6724239" /translation="MGALISYAAPCGALGGAEERGQGDAGALAPAREMAAEVA AGQPGAVMLVARKDDVHDVAVGTLEGSAPMRDITIFIASPTKAVTAYAMVLE EGKIDLDSPVDRMLPELANRVLARIDGPIDEYVPAERPIVRDLMTFTGFSFDA SSPQRAIDELGIVNAQPVPTPHGPDEWIRLGTLPILMHGPQOMMYNTGSIYGVVL VGRADOGDFAEVFERILAPLGMRDTFHPADKLAFACGGYTTDGTGKTTMDHD GAESAYSAPAPSPGAGIVSTVDYDILFARMIMGVHGGRLLSAVSREMTADHL TPKAEASASPEPGEFTFHGMCTGMAVTTAPDANVSEYGRIGMDGGFTSWINDGREL IGIYMTOSAGLFEGALEFRMSYVATESA" 3415..5556 /note="ORF3: similar to Na/H and drug/H antiporters" /codon_start=1 /transl_table=1 /product="unknown" /protein_id="AAF26906.1" /db_xref="GI:6724240" /translation="MHGITEROVLLSVTLALILVTARASGELARLRLOPEVIGELFG GVTVSGVALAGPGRHALPEBPVAVCVLSGIMTALLMAGLEVGVGILRKKA RPGKLSAGIAPPLAGAAFSALVDRPIPSGLFGLIVSYVRSYIAVILIEREM RRSTAOVTLAAGVSEVAAMVLAMTSSSTGASPLAANVASALASFLVLVAVGRK LTHLRMVAADATRVSKGSQVSLVLTFLFGLRGLRGLGAFLLGVLNSAPTR NRLPLDGVQTLVAGLFAPVEFVLAGMVDVSLQTPPAAMGTVALITATAAKVPA LGAELTDRGSEALVAVGLNMKGDTLIVAVGELTILSNEATYMWAAVALVTVA ISVALITLKRAPPTOESARLEEREAARAYIPGVERRIIVPIVAHLLPDPADIVS IVASKRLGETVDTTELSTVEQAPGSPRAGAESRGILARGRLRVCIMORREIRKS IQALIRSRDHLVLIGARSPARAGSFGRLDQATVQKESNVLVVVGDPAAERAS ARRVLIVIGLEYSFAADLAHAHVALMDELYLSSAQDPPGAVVWRDEPSSVRA ARSLVDAEVRGRLLGVRSVSVHVGAMPDELTRELARAPYDLVLVGCYDHGPLGRL YLGTEGVSVVRSVPALLVYAHGCTREQR" complement(5612..5592) /note="ORF4" /codon_start=1 /transl_table=1
CDS	

```

CDS
    /product="unknown"
    /protein_id="AA26907.1"
    /db_xref="gi:6724241"
    /translation="MDKPIGTGRCALAGGYIPGSGNGPEPOMTSHETACLLNASORDA
    OVAITYFSDRDPAQPRVYTPARTRTRHVRFDLLEPERIPEDTQYASVIESDAIV
    OHTRLDSQAEALSTIAYTDRE"
    6236..6675
    /note="ORF5"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AA26908.1"
    /db_xref="gi:6724242"
    /translation="MKHVDGRRFRGRIHTLGLLASMALACGGSPSEKTVQGTPLAP
    GADAVTADVPDPAATRLAVDVHLSPPERLEAGSERPVVQRPSPSPMWRVGVLD
    YNADSRGKLAETVYANFELLITAEKQSSQSPSSAAVIGTSVG"
    7610..11875
    /gene="epoa"
    /note="epoa"
    /note="EPOS A"
    /codon_start=1
    /transl_table=11
    /product="polyketide synthase"
    /protein_id="AA26919.1"
    /db_xref="gi:6724253"
    /translation="MADRIERAADPIAIVGASGRLPGVITLISGFWTLLESGRDV
    GVPAERMDAAMFDDPDPAQGTPTVTRASFSDVACDASPFGISPRALMDPAHR
    LLEVCWALEMAAIAPASALVGTETGVFGIGSPSEYEAALPQATASRIIDANGGCTM
    PSVAGRISTYALGLRGPVAYDTATSSSLVANHLCOSIRSECGSTALAGYSMLSP
    STLVLSKTRALARDCRKAESADGFGKGGCAVVVKRLSGARAGDRIILAYIRG
    SAINHGSASGTLTPNGSSOEIVLRLADAGCAASVGVYEHAGTGTLLGPIDIOA
    LNAVYGLRDVATPLLIGSVKNTLGHPEASGITGLKVLVLSLQHQIIPAHLQAOLN
    PRISMGDRLTVTTRARTPMPDNTNPRAGVSFGSGTNNAHVLEAPAACTPPAPE
    RPALLVLSARTASALDAQAARLDHLETPSQCCDVAFSLATRSAMEHRLAAT
    SREGRLALDAAGCQTSFGAVRSTADSSRGKLAFLTGQAGQTTIGMGEGYLDVNSAF
    REAFDLCVRLFNOELDRPLREYMAEPASVDALLDQTAFTQPALTEGTLDAVMSR
    MGPELVAHGSIGELVACVAGVPSLEDVAVLVARGLMALPAGGAMSIEMPEA
    DVAANVAPHAASVSIANAAPDOVITAGGQVHAIAAAMARGARTALHVSIAFHS
    PLMAPLMEFGFVRSVSVRPSIYVLSNLSGKACTDVSFSGVYVRRHARVVRADG
    VKALHAGACGTFVVEGPKSTLLGLVPAKMPDARPLLIASSRAGREPRATVEALGIM
    AVGLVSNAGLFPSCGRVPLTTPWQREKRWIDTKADDAAGDRRAGAGDEVEEG
    GAVRGDRRSARLDHPPESGRERKEVEAAGDPRLEIDEPEVLDHLVATERAPG
    LGEVIAVDAGLSFNDYDOLAGMVPDDLPGKPNPLLIGGCACRIAYAVGVNGLV
    VGPVATLSAGAFATHTTSAALVLRQALSAIPAAMPVAYLGVAVYLDRIALOP
    GERVLIIHATGCVGAAYOMQHGAVHATAGPEKRAVLESIGAVYVSDSRPFI
    ADVRAWTGCEVDYVNSLISGELIDKSNILRSHRFPYELGRDCAANOGLRPFIR
    NLFSVLDRGMMLERPARVRALIELELGLIAGVFTPPITATLVARADFRMAQ
    AQHLKLVLTGDDPEVQIRIPTPHAGAGSPSTGDRDLDRLASAAPARAAAEAFRLTQ
    VSOVLRTEIKVGAELFTRLGMDSLMAVELENRITEASIKLSTTFLSTSPNIALA
    ONILDALATLSLEVAENLRAGVQNDPVSAGADQWEIIL"
    7640..8920
    /gene="epoa"
    /note="Region: beta-ketoacyl synthase/malonyl-CoA
    decarboxylase"
    9236..10201
    /gene="epoa"
    /note="Region: acetate acyltransferase"
    10529..11428
    /gene="epoa"
    /note="Region: enoyl reductase"
    11549..11764
    /gene="epoa"
    /note="Region: acyl carrier protein"
    11872..16104
    /gene="epop"
    11872..16104
    /gene="epop"
    /note="EPOS P: contains condensation/heterocyclization,
    cysteine adenylation, putative NAD(P)H oxidase, and
    peptidyl carrier domains"
    /codon_start=1
    /transl_table=11

```

```

    /product="nonribosomal peptide synthetase"
    /protein_id="AA26925.1"
    /db_xref="gi:6724259"
    /translation="MTINOLNEHEHOGIKILADGERLOIOAPKNALPNLLARISEH
    KSTILTMLOKLPASIVPAPAEHRAPPLDIDGESVLTGRTGAFTVPSGTHAVREVD
    CTDLDPRLSRFRKRYVARHDMRLAHTLPDMQVTEPVDVDEIDILDRGDRSTREA
    RLVSURDMSHRTIDYTERPPLYHVAVALRDERQTRVLVSLINDLGSLSITIRKDL
    SEYEDPETSLEPYLESYRDVYLALESRSKAHOSMDYMRRLAELEPPPTLPMKAD
    PSTLKEIRFRHTEQMLPSDSWGRLKRGEGGLPPTGVILAFSEVIGRMSAPFTL
    NITTELRPLPHVRVNDITGDFTSMVLIDITTRKPSFGOARAKRIOEDMADCDVS
    GIEVOREARVILGORGALFPVVLTSALNOQVGTSLORGETPYVYSTOTPOLLOH
    OLYEHQDPLVLANMDVDCVFPFDLDDMLAEVYVTLRLTTEPMEGEOYRCSLPQOLE
    ARASANATNALISEHTLGLFRAREQPLQDLAVVSAKTLTTELSRRSRRLARLE
    EOGAPNTLVAVMEKGEQVAVYLALESAAVYPIDADIPARITHLIDHGEVKIL
    LTQPLDLKLSMPGIGORLVSSEGVESGDGQPPMPPIOTSDALVYVYTGSGTGLPK
    GVMIDHRCGAVNTIIDINREFEIGPDRLVLSSEFDSVDFGIIAAGSTIVPDA
    SKLRDPAHMAELIEREKYVNSVPAALRMVLEHREGRDVLASIRLSISGWMIPY
    GLPRLQAIRGVSYISLGAATEASINSIGTPVRVNDLSMASITPGRILRQTVLD
    EALERPVPVWPGQLYTIGVGLALGVMRDEETRSFLVHPETGERIKYKTGDLGRYLPD
    GNIEPMGREDNOIKLGRYRVELGEIETELKSHPNVRDAVIVVGNDAANKLLAVVVP
    EGTERRAAEDDASLKTERIDARAHAAEADGSLDGRVQFKLARHGLRDLDGKPVVDL
    TGODPREAGLDVYARRRSVTFLEAPIFVEFEGRLSCLSSVEPDGATLPREFYPSAG
    STYPQVQYAVYKSGRIEGVDGEGFYVHPEFHEHRLKLSHGIERGAHVRONDVDEEA
    PNLFVGRIDALIESLYGSSRRECIIEGAYALIMEGAPSCNTGVCPGVGFNPEQYR
    PVLDLRHSDVYVHGMIGGRVDPQVCTTLODSSPRATTGAPGREQGFADMLRD
    FLKRLIPEYMWYFVELDALPLTNSNGVDKALREKRTDSSPRHSHTAPRDALIEI
    LVAAYREVILGEVYGLQOSFYDLGATSTHIVRMSLSLQKRIDREIATILELRYQPNLGS
    LASGVRDSRDLDPNMDRVEYRRKRRRS"
    16251..21749
    /gene="epob"
    /note="epob"
    /note="EPOS B"
    /codon_start=1
    /transl_table=11
    /product="polyketide synthase"
    /protein_id="AA26920.1"
    /db_xref="gi:6724254"

```

```

Query Match      100.0%  Score 101; DB 1; Length 68750;
Best Local Similarity 100.0%  Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCGGCTCGAGGCGCTGCTCCCGACGACGAGCGTTCGCGCCGACCGCAGC 60
    |||||||
DB 37682 AGCGGCTCGAGGCGCTGCTCCCGACGACGAGCGTTCGCGCCGACCGCAGC 3741
    |||||||

QY 61 GCGGTGGGTGGAGCGAAGCGTGGCCATGCTCTGCTCA 101
    |||||||
DB 37742 GCGGTGGGTGGAGCGAAGCGTGGCCATGCTCTGCTCA 37782
    |||||||

RESULT 3
ARI93029 LOCUS ARI93029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

```

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 37741
|||||

OY 61 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 101
|||||
Db 37742 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 37782
|||||

RESULT 4
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 37741
|||||

OY 61 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 101
|||||
Db 37742 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 37782
|||||

RESULT 5
ARI99559
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 37741
|||||

OY 61 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 101
|||||
Db 37742 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 37782
|||||

|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 37741
|||||

OY 61 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 101
|||||
Db 37742 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 37782
|||||

RESULT 6
ARI99567
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 37741
|||||

OY 61 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 101
|||||
Db 37742 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 37782
|||||

RESULT 7
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 68750)
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 8.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 60
|||||
Db 37682 AGCCGGCTGCGAGGCTGCTCCGACGAGCGTTCCTCGCGCAGCCGAC 37741
|||||

OY 61 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 101
|||||
Db 37742 GCGGTGGGTGAGAGGCTGCGCCATGCTCTCTCTCAA 37782
|||||

Db 37742 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 37782

RESULT 8
 AR208671 AR208671 68750 bp DNA linear PAT 20-JUN-2002
 DEFINITION Sequence 1 from patent US 6383787.
 ACCESSION AR208671
 VERSION AR208671.1 GI:21509886
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 68750)
 AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
 TITLE Genes for the biosynthesis of epothilones
 JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
 FEATURES
 source Location/Qualifiers
 1. 68750
 /organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
 ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
 Best Local Similarity 100.0%; Pred. No. 8.8e-14;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCGGCTGGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTTCGCGCGAGCCGAC 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 37682 AGCGGCTGGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTTCGCGCGAGCCGAC 37741

Qy 61 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 101
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 37742 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 37782

RESULT 9
 AR172664 AR172664 71989 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 2 from patent US 6303342.
 ACCESSION AR172664
 VERSION AR172664.1 GI:17912155
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 71989)
 AUTHORS Jullien,B., Katz,L., Khosla,C. and Tang,L.
 TITLE Recombinant methods and materials for producing epothilones C and D
 JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
 FEATURES
 source Location/Qualifiers
 1. 71989
 /organism="unknown"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
 ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
 Best Local Similarity 100.0%; Pred. No. 8.7e-14;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCGGCTGGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTTCGCGCGAGCCGAC 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 32070 AGCGGCTGGAGGCTGCTCCGACGAGCGGTGCAAGAGCTTTCGCGCGAGCCGAC 32129

Qy 61 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 101
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 32130 GGCCTGGGTGAGCGAGGCTGCCCATCTCTCTCTCAA 32170

RESULT 10
 AB070940/c AB070940 104326 bp DNA linear BCT 22-JAN-2002
 LOCUS

DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.
 ACCESSION AB070940
 VERSION AB070940.1 GI:15823967
 KEYWORDS
 SOURCE Streptomyces avermitilis DNA.
 ORGANISM Streptomyces avermitilis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.
 REFERENCE 1
 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
 Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osone,T.,
 Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
 TITLE Genome sequence of an industrial microorganism Streptomyces
 avermitilis: deducing the ability of producing secondary
 metabolites
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
 MEDLINE 21477403
 REFERENCE 2 (bases 1 to 104326)
 AUTHORS Ikeda,H.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
 Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
 Shirokane, Minato-ku, Tokyo 108-8641, Japan
 (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
 Fax: +81-3-3444-6197)
 FEATURES
 source Location/Qualifiers
 1. 104326
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 complement(1..1134)
 /gene="mcma"
 complement(1..1134)
 /note="mcma"
 /note="methylmalonyl-CoA mutase"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB69185.1"
 /db_xref="GI:15823968"
 /translation="MTGYSTADANAREFQVSGTRAAPVYFDLPTBAGLSDAPAA
 RGLVGRGVAVDSIDDMVLEFGELDLVDSVSLVADAPALVLYOVLAEHGAAPR
 RLAVGRGVDFEKEFLARGVPELPSPSLRMNVPVACQALPQVCSVGHNLAC
 GADPLAEVAFYVAGIEYLRAAVYAGLVRYVAPPLTCLVGGTGGRRKKLRTR
 VWTIRANEFSGSRTAPRLVYTPPRDGLATCTEDTADGATVADLRIERLGGTLTVL
 EHGILAGQLGRPHGDRPATPRPHPHQSDLDLQAROTERLAKLRMRVQPAVDALLR
 LRKAQGDNDVLYPMREALLAARATVGEVCDALRSTWGTGPSAT"
 complement(1131..1886)
 /gene="ppia"
 complement(1131..1886)
 /gene="ppta"
 /codon_start=1
 /transl_table=11
 /product="phosphopantetheinyl transferase"
 /protein_id="BAB69186.1"
 /db_xref="GI:15823969"
 /translation="WRAHEVFDVPSVSLHPREALVODAVESRRRRFATGRHCARRA
 LASLGVPGPLATPRGAPRWPAGVCAITHCAGRAAARAPGIRAVGIDAEPRP
 LPBGVLAVALPAEOFOYKEHTARHPGVWMLFSAKESYKAFPGATGPRLEFEA
 LIDLDVPSSTFVRVLPALRAAPELDLDLGRMLVIRGILVTAVVYDGPDRDPNAP
 SPYDKTKSAPTRGSADRGCRPTDRATDRFA"
 complement(1910..2875)
 /gene="simx4 homolog"
 complement(1910..2875)
 /gene="simx4 homolog"
 /codon_start=1
 /transl_table=11
 /product="simx4 homolog"
 /protein_id="BAB69187.1"
 /db_xref="GI:15823970"
 /translation="MSODRHDSSTAALGSPSHGRLAVSDIHYRHRRNDIYVDLHP
 GCAEDMLIVAGDVASVADITWALSILBDRRAKYVIMAGNHELTARCDPDLGERK
 YRHLVDLCRALGVYTPEDPTFYWHGAGCPVAVAFLLYDTSFLPRGMSKSHLRILA
 DDAGVICTDEFELHPDPYPSREAWCARAVRATRALDLPDLPTVLVNHPLVREPT
 RVLRRPEPALWCSTTATADWHLRYRAAIVVGHILHPIPTIRCDVDPHHEVSLGYPREV

gene
3213. .3971
/gene="olmc"
3213. .3971
/gene="olmc"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="Bab69188.1"
/db_xref="GI:15823971.1"
/translation="MSANRYDGNLMWRQFHPAPTAATKIVLCLEPHAGSASYFLPVSKA
LAPGVYLAIOYPRQDRIHEKCIDNIAELADAVTALEIADRPVLEFHSAGATLA
FEVALRIERQIVPLAFSGRAPSRHREDSYHLRQDDEIIAELKTLSTRESEVLD
EELIRMYLPAIRSDYKAETRYVAGGARLTAPIHAHGVDEPRVSLDEAQAAMDHITG
HELRSHGPHFLINDQAPVIEIARVSSVAS"
4643. .7438
/gene="olmri"
4643. .7438
/gene="olmri"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="Bab69189.1"
/db_xref="GI:15823972"
/translation="MLIGRGOEIDRLRAAVSVSTGRSAVLVESAGGKSEILED
LVHVESIAGILARATYIAERNTPLSVLSVNGAPPPDDLVRFPHMDDAARMTS
TEGRDKARRDAANYELMRTCTEYRALAQRAPIVAVDDIHGDAISRLIHLARHC
RNVPLITFTQSLAGGPADTFEALRQNLHRIPLRCRODIDMLMEQSGADG
QGIYVERFAVAGSGNPILLRALLEBDARLEPDVDQADVGPAQAYGRAVLACLRGG
EKAEHYEAGLAVGCAFTHELRLDLSAALTEMLRALAGIYVGSFOHPAAK
AVADSLERSHRLRHORSALLHAGSPALAEHLHEAGADGAMAVGYLOOAEQA
LADBARAVSYLEIAAASADPRGREIKRIGSVTRYVSADAEHVDALRGGA
GLTPEPTGGLAALLAHGRLBOARSYLEGRRRGCGDARTARRLAIIROSVLMR
GANCAGTRAAHNPGRSTGAGTGLRAMLSGSHALLAQRIAGTEGLLETHLD
ATPDVLNVAWTLALGAVERALHWCOTLEAKRDAPEGMEAFAAVFAEGALRGD
LABERDARCFALVPERHRSALGSAVSYLVQIARKEYEDAROLSPMAISLPD
TTHMLDYLRARGLYMATROYHSALRDEHVGRLAERMGDPALLPMTDTAEALR
LGREEOGLRLATEOLIAVODGAPRIKISIRLQGLSEPAARGLILRAVAVELSRAD
RLTSLRTPLDALAHRETROTROAAMYLRRAROLAKSCGQPPDRIGCEPDAGASGE
LREBGPERNADMGHLSSESRVAVGASGTTNREISVRLITIMSTVEQELTRYRKLK
INREELLAERAGVOY"
7658. .10486
/gene="olmrii"
7658. .10486
/gene="olmrii"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="Bab69190.1"
/db_xref="GI:15823973"
/translation="MLASAGQGBGVVLYVNGPQVGTALDLAFKRAITSGARVCA
AGSSAETEAETFGVGLPPTGPIGAGFRLARATIGSDAGPGLDADRLFALCOELER
RPVYIAVDAQIOLADASLRFRLRYLRRLRAPVMVLTPTGALPSPQAEILRRP
QCRRLGILPLSRDGVQVYSAQVSDVRPLADFNHISGNDPLVHALIADHRAAR
LSBPVSVPYVSSASASASASASASAAVAAAPAAFKKAAIAMVYRPEVYF
DVAGCFVILGESATPAVACTIVDGGADAVOVMTADTALLGQGRSLVQALILE
HUVETRAELRRRAARLHGDCAPATVVAHHILAAAPAEPMARITLILAAEKMOBER
EELSLDCLRLAGSATTBOERAAYVAAKVICMEIDRIVSPWIDELVYALRKDLS
GDAAMIYKHLAMQDREPEAVESLSTLAAADAGEPVLELHVRHMLHNSCPLENS
VPGSNPTRPRTPLRLRRVSPASVAGEVLVLEGGPDHVAABEELINCRSETVY
EAEALALALVAGRPDRAFCMDGLLOQAGTSGTGSVALISGIRAEILRGALPA
VEOYARSALSAISLPGWVAGVSPILALVHSTTASGSGVAGAMINQAVDGMERTH
GLYTHARGHYHLAVDRPAALDDDLTGGLAKEMGDVPTCLPMRTSAOQALCN
PGQARALVREOLGRPGGSTRAKGVALLRLAATGEPERRALLRESVNIJESGDOYE
LRLALDRCOLDHVGAAPAKARMTARARNAVADCGDALFRKLFREVDREDPDVG
LTGEEDGSASLTAERVRVTLAALGHSNREIGRLFTTKSTVEDHLTRYVRKLGVRNR
ADLGDLAGTNLGAGTPAMAMASSAVAG"
10933. .11271
/codon_start=1
/transl_table=11
/product="P450-like protein"
/protein_id="Bab69191.1"
/db_xref="GI:15823974"

gene
11458. .29898
/gene="olm1"
11458. .29898
/gene="olm1"
/codon_start=1
/transl_table=11
/product="modular polyketide synthase"
/protein_id="Bab69192.1"
/db_xref="GI:15823975"
/translation="MACRLPGAAPFEALLRNGEDAVGAPGDRRYARRGFLDS
VDFEDGREGISPREAAMPQOORLIVLEISWEALDEADGIPPALGGSGSVYGAIALD
DYATLSRAAGADATAPOTTGTLKRGITIANRVSTTLGRGSPFTVGGSSSYVAVALA
AESLRKECSVALAGVNLTLAEDSTVAERFALSPDGTCTFDANANGTYAGEGG
VVVLEPLADAIDAGDDIICILAGSAVNNDGGGLTPVPDAGQAVLTAVEQGISP
AAVGYELHGTGTPTGDPVEAAAVGAVLGAAGPACOPLVGSKYKNIHLEGAATAG
LKAVALVHREIIPATLHNSPNAIPMAELIGVNTALPMPEDAGPLVGVSSFGM
GGRNCHVYLTEPAGYDAAAPATAGIRPAGTPAPLIVSGDEQALRAOARKLHDLT
HGBLPLDVAHTIAAARAPLARAVILLADREVSAGRLANDGTCGADLVRCVSPG
ALAFVFTGQSSQRPMTAELRTSAETAAALDEVCSLDLWMLRPLRDVLFABEGDE
ALALDRTFTQPALFAVEVALFRTABEHGCTLPPLFLGHSVGEILAAHVAGVSLADAC
TLVAARGRLMODRPATGAMYSVQATDEELAPFVSESVTLAALAGVITAGDEDAVL
AAVAHNAARGKTRKRLRSHAFSPHMDGLTFERRIAGLTFGAPRIPIVSNETGL
LTEPADRSPEYVRAAARVAFIDGVRLLEDQVTLTLELGPVGSVAMARDCLRAG
TYCVPLLRGDRTPERPALIALGLVRYGVTPMERTITBAGARRPLTYTANQRHH
WLPEPSASAPRVVNGGAGHQAQGHQAQRAOQAHAGDGPAPRAVNRDTPVELYR
DRDTLLEPDAATQVILGASHPDVATVHEARELFGFSLMAIEGELIATYAGRVPT
LRLDHLTPDRLARRLEADGAEARTDAGPVPAAAHDDPITVYAGKGRPLGCVRSPE
OLMRPVAAGDALISALPVDGMEFADFPCEGFLADVAFEDGFFGISPRELAMPD
QORLLTSEMEALERAGVDAIKLRGSRGVFAGVSGSEGRPLHETSDDGHYLTGTA
PSYLSGRIALVILDEGRPALIVDTACSSYVALILIAOLRGCSDIALAGVYVMTS
GNFETRAKQGLAPDGRCKFAADGAGTGEBGVMLAYORLSDAYDRGPVLYAVRG
SAVNSOGASGLIAPRPSQOARIKALASGASADVDVNEAHGTSIGDITEQA
LMTYGGDRDDDDPLGMEYSKNSIGHTQAAAGVAGYIKSVLALRNGLRPTLHVDP
PEVMSAGAEVLETPGEMETPEGPRRAGSAGISGTNAHVILTEQAPSVEEVTA
DDGLVPMWYASBEALROARLRADHVESGLPADVGLSLAVRAGLEHRAVVA
EDBAGTAHLIDALBERTPAGVBEGTALFEGGCSORAGMGRRELYSAHPVANAAD
EVCARDGLILPGLRLEVYVEGALIDRTEMAOPALRYALVAFELVSSKGYRADVILGH
SIGELAAVYAGVSLADACRIYAAKGRLMQALPEEGAMAVFAADELDLPFGVSV
AAVNGPSRLVLSGDEBPVTLAQAFABOGRRTQOLYSHAFAHSAEMPIAEFETLA
AVEFRPRLIPVNSVNGEMAGDEFTTPAYVVRVRAVVRADCGVAAHAGVRYTEL
GPDGTLTTLAHDPTAHAGRDVAAPALRGQPTFRVYVTVAGLYTGLTVDPAFEA
PYCARPELPTVAPOHRRWYLEPVAASAPRAPLRYTAAVSPGSPRSISGTAAV
VVPASGSEYDITVAGTVALRDHGCARPVYVVEVADGEPERADVATLRAALGEDMGVYS
LLGTAGSLSSGVATATVLVQAWMTDLGGGARLWCLTRGAVSVPDRLAARQOLMGW
GRVAALHEPPEMCGVLDLPVAPDDRMTRLCAVLAAGSGEDQYAVRAAGIFARRLGRA
GTRADAPAPSPMRDGTGLVYVGGTALCAHLAOWMLAAAGAAHVLLTSTRGPDAGA
AELTARLRETGTETVAACVPTRDALAEILSGLPADRPVLTGVYHAAGVYLDGVLSD
TPRPRPAAVAGPKYTGAMHHELTRODLSAFVYFSSFAGVYGLAGQANAAAAVYIDA
LAVHRQDGLPATVAVMGSWSGAMADVTEAAROLKRTSLVPLADALALAGOVND
DGBTAIVVADVDWDRFAAGFGPSPSLTGLVPEVNRIEERFERGRGAAPETDRDLGL

Query Match 59.2%; Score 59.8; DB 1; Length 104326;
Best Local Similarity 76.8%; Pred. No. 0.00016;
Matches 73; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 4 CGGCTGCGAGGCTCGGCTCCCGACGAGCGTGCAGAGCTTCTCGCCCGACCGCAGCGC 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100340 CGGACGCGCGCGCGCCCGCGGACGGAGGTGCTCGCGCGCGCGCGCGGCGGT 100281
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 64 GTGGGCGTGGAGCGGAAGCGTGGCCATGCTCTGCT 98
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100280 GTCCGCTGGGCGAGGCTCGCGCTGCTGCTGCT 100246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
AF411573S4 3185 bp DNA linear BCT 27-SEP-2001
LOCUS AF411573S4
DEFINITION Actinomadura verrucosospora polyketide synthase 2 gene, complete
cds.
ACCESSION AF411576

VERSION AF411576.1 GI:15788318
KEYWORDS 4 of 4
SEGMENT Actinomadura verrucosospora.
SOURCE Actinomadura verrucosospora
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptosporanginae; Thermomonosporaceae; Actinomadura.

REFERENCE 1 (bases 1 to 3185)
AUTHORS Tsantrizos, Y.S. and Yang, X.
TITLE Cloning and characterization of a gene cluster involved in the biosynthesis of veruocopeptin, a cyclodepsipeptide metabolite of Actinomadura verrucosospora

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3185)
AUTHORS Tsantrizos, Y.S. and Yang, X.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-2001) Department of Chemistry, Boehringer Ingelheim (Canada), Laval, Quebec H7S 2G5, Canada

FEATURES
source
1. .3185
/organism="Actinomadura verrucosospora"
/strain="0886-2"
/db_xref="taxon:46165"
1. .3126
/codon_start=1
/transl_table=11
/product="polyketide synthase 2"
/protein_id="AAI07759.1"
/db_xref="GI:15788322"

CDS
/translation="MDADADRAADGADRAVERALRALLEERDLRRRTDILKAGGEP
IAVVMACRFPGVSTPEGLMEAVRGGRIVEEPDRDWRDVPDEDPGLTCTSTRH
GSELTGAGFDFEFEGISPREALAMDPOORLLLEASMEVEFAGIYVADVGTAGVF
GSSVPEXGLRFLFEDOHLEHGLSLGSLASVAGRAVQLGTGPAMTVDAAGSSSL
VALHIAVSLRSGECSLALAGAAVMTTPAYVEFSRQGLAPDRCARAFNAADCTG
WSEGVILLERLSDQRNGHEVLAVLKSAVNQDASRGTLAPSGRQEEVIRKALA
DAGIASEVDLVEAHGTGTLDPIEAHALMTYQGREDRPLWLSKSLYKHAQA
AAGVAGVIAKAVNALRHGYLPKTLHDAPSRYHMGCAVEELLVDREREMERGPRAG
VSSFGSGTNAHVILIEAPRKGAALPEIPGLADGLVPMVAVSGKPAALAOAALR
DEVTAVSGADADADIGMSVSRSAFAHRAVVLGRDRDELLSLASEGAEAGV
RGVAGELGVMPFGGSGOMAGMGRRLYDAVPVFAHALDECAALDERAGSLDVI
RGEAGAPLERDVIQPALFSVMSMALMRWVEPAVAVHSGEILAAHVCALSL
LRDAKVAALRSRALVDLAGHGMASVAESADVAEOLAPMGRLGIAVNPGRSVV
SGALDADLEFLERKMAEGVARARIQVDAISHVEVARVEQVIGPLSDGPTSLPF
YSTLEGEVIDTAELEDGDWYTNLRKVLFESSGRDLADGAVFIEMSPHPLVTVPO
EMVEDGAVVILSSRRDGEVAMADSLAQLVVRGTVDMALFETRRRTDLPYAFQ
RRRYLNSHTRAVAMELPADAAEEDQVTPLEPVVALPEGDALVLDHYLEKIAT
VLGHAPGESVDPDRFETKDLGPDSTLAMDSLRLASTGTGRLKPTLVLRHPSRLIAGH
IAAVMLDRRPA

BASE COUNT 395 a 1086 c 1254 g 450 t

ORIGIN

Query Match 58.6%; Score 59.2; DB 1; Length 3185;
Best Local Similarity 76.0%; Pred. No. 0.00047;
Matches 73; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 3 CCGGCTCGAGAGCGTGGTCCCGACGAGCGTACAGACTTCTCGCGGACCGACGCG 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 762 CCGGCGAGCGCGGCTGGCGCCCGACGCGCGGTCCAAAGCGGCTTCCGCGACGCG 821
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 63 CGTGGGGTGGAGCAAGCTGCGCCATGCTCTGCT 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 822 GACCGGCTGCTCCGAGGCGGCTGCTGCTGCT 857
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AF016585 41097 bp DNA linear BCT 07-DEC-1997
LOCUS AF016585
DEFINITION Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)
gene, partial cds; polyketide synthase modules 1 through 7 (nidi)
genes, complete cds; and N-methyltransferase homolog gene, partial
cds.
ACCESSION AF016585

VERSION AF016585.1 GI:2558836
KEYWORDS Streptomyces caelestis.
SOURCE Streptomyces caelestis
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 41097)
AUTHORS Kakavas, S.J., Katz, L., and Stassi, D.
TITLE Identification and characterization of the niddamycin polyketide synthase genes from Streptomyces caelestis
JOURNAL J. Bacteriol. 179 (23), 7515-7522 (1997)
MEDLINE 98053867
PUBMED 9393718

REFERENCE 2 (bases 1 to 41097)
AUTHORS Kakavas, S. and Stassi, D.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1997) 47P, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064, USA

FEATURES
source
1. .41097
/organism="Streptomyces caelestis"
/strain="NRRL-2821"
/db_xref="taxon:36816"
complement(<1. .396)
/gene="nidi"
complement(<1. .396)
/gene="nidi"
/note="ORF6; similar to tly1"
/codon_start=1
/transl_table=11
/product="cytochrome P-450 hydroxylase homolog"
/protein_id="AAC46023.1"
/db_xref="GI:2558837"
/translation="MVDSTGGMELSKANAKELLDMFSNRTTHHPVWDGROAMOV
FRYDYLTVSNHPERFSSDFTEVATPELEMLIGRTIGALDPPAHOPMKLYSQAT
TPRRAGGEQRIRVIAEELDRVKQKTTA"
627. .13649
/gene="nidi1"
627. .13649
/note="nidi1"
/gene="nidi1"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 1 and 2"
/protein_id="AAC46024.1"
/db_xref="GI:2558838"
/translation="MAGGDATAQKAQDAEKSDEGSDAIINIGSCRRPGAPGAFFN
OLSSGADAVVTAADGRRRTTIDAPADDAFEGNSPREAATDPQORLVLELWEAL
EDAGIVPELSLGEAASVFGAMNDYATLHRAGAPDTYTATGLOHSMIANRLSTFL
GLRGPSLVVDTGOSSLVAVALAVESLKGTSGLAAGVNLVLAEEGSAEERVGL
SPDGRCHTFDARANGYVREGGAIYVLRPLADALADGRRVYCVVGVATGNDGGPGL
TYPDVGQGEAVILRAACDQAGVPADRVVEVTEHGTTGTPAGDVEAVALGAVGTGRPN
EPLVGSYKTNIGHLEGAAGIAGVPKALCCHERLPSLNEFTPNPAIPIERIRLKY
OTAHAAIDPVTGGEPFLAGVSATFGMGVNCVLEETPGKQPEGTGADCLTSASP
MLLLSARSEQALRAQARLRHLEDSGADPLDIAYSLATYTRREHRAAVCGPDRLL
SSALAAALAQOTPRVIRIGSTDAGRLALFTGGAGQHPGMOELTYTDPHRAALDE
VCEELORCGTNRLEVMETPOPLDTEETTOGALFLQALVRLAEHRAEDTOAHVLY
GHSVEITAAHAGVLDLPDAARLITAAHMGOLPHGALMSQAQAEHDLDAHMH
GVEIAAVNGPCHVLSGRTALBETAOHLRQONRHTMYKSHAFHSLMDPMIGAR
DTLNTLNTPTPLILSNLTCGLADPNLCTPDVIDARHTVFAADVQAHNGT
TYLEIGHPTLTLLHLLDNPPTLPTLHRRPEPELTQIAAAGVATDIDAVALC
GASRRVLELPTVAFQRTHMAPGLTPHMAADGAPPAEPQAMVAGVSRBALRVLG
ETTASVLIDGDEVALDRPTSGOLDSTAVDELGLGTAGVALDPTLYVELPTPR
AVADHLAKTTLIGESADADQDNGTNGTGAEEKAGDPIAVIGICRPFQGVATPDIME
LVAAGTDAISTFFPDRCDDGIDYDPDSTGKGYVRRGGTLHNAOPDAFREGISR
EATADDPQORLLETSWELERAGVPELSLGGRGVFGVTTAEYEGRHEGDTGE
GELLTGTAASVAGRIAYALGTGRLALVDACSSSLVALHAAVQSLRSGCDALAG
GTTWASGGMFEESRQGLAPDGRCKAFSADADGTAAEAGVMILVERLSDARLGH
RVLAVERCTAVNODASNGLTPAGSQAQOYIRDALSDAGSADDIDAVEAHEGCTGL
GDLPEAGALATATYGHPRQTPVWIGSLKSNIGHQAAAGIAGTIKMQOALHDHLLPT
LHADHPSKYVMDAGPIOLLIDARPNPADPDRPRRAGISAGVSGTNAHYVLEPPAD
EEAPRTQAPPPVIAWPLSAHTPTALRQARLRDLSMLHTHTDADQDIDAHALATRTHE
KRAVITGRTRELTHTKLHTLDAIOGTAAHPRHRLTFTGGAOHRMGQBLVATDPIH

FAALDEVCELORCSTONIREVMTFPDOLLDRREYTOPALFALOTALYRTLTARG
TOAHVLGHVSEITTAHITAGVLDLPDAALITRAHMLGOLPHGGMALSVQAAEHDL
DOLAHNGSVETIAVNGPTCHVLSGSPRLALEETAOOLHOGIIRHTMLKVSALSMALD
PMGAFRDTINTANTOPPTPLISNLITGOIADPNHLCSTPDYWDHARHYRPAAYOT
ADQRTTTLTETGAHPTLTTLHHTLNDPTTLTHERPEPELTTLTALTLHTTGTT
TTLHTTSPQTHHDLPTYPQORDRYWMEPRVAQSGOPADRLRYRWMEAPEDHT
VPAERMLLGGSDPADAGILVEBOLVGHGAAYORLIDIPAGRPLISCESELPR
GSHTEFCGSRSPRAEVALLOAGAGAVORLIDIPAGRPLISCESELPR
OLMGIOVAGLELADRMGGLVDLPADAPALRGILARLVSNAPDNQVIRASGVPR
RYVGRPRYRNDMAPSGYLTITGCGALSGOVARRLALAGAPHLILAGRGSGAEP
LVDELALGAEVTVACDADADRALVSLATPEHRITLVAAGLVLDGVLGILTP
ERIDAVLAKATARHLDLTDADLDAEVLFSIVGWNGGOATYAANAALDALA
HGRARGORATSIANGWAGSGAAGAKSLARDGIDVLEPEALDLEOARSET
TLIVADMETFLSSASRRFOPLFCRIPARAVANGADSPDVPVLSLALHVAQSLA
EROLRLIDIVTEASAVLRHGSATVADEAFARAVGSDSLTVELIRRLASAGIRLS
NTLPRPPTAAAYELBOLCTESBAASHODACARTYVADERYAVVGMACRIPG
GVATPDDELVASGTDALISTFPDRDMDGLIDPDPSTPGKSYVHGGFLIDAAOF
DAEPFISPREATAMPQORLLETSMELHAGIAPHTLTTRTGTGVTASQOEEGT
OSREADAKYGGHLLTGLTASVMSGRAVYTLGLOGPALVDTACSSILVALHVAQSLR
RGCDLALAGSTVMTSPVVEFESROGLAPGRCKPFADTADGTAMAGVGLVVE
RLSDERLGHVLAAYRGTAQNDGASNGTLAPGPAQOQVITDALSDAGLSADDDA
VEAHGCTPLGDPTEABALATTGHEROPTVWLSKSNIGSQAAGACITIKMO
AMHGTPLRTLHVADPTSOVMEAGSLQILAEARPWADPRRAGISAFVSGTNA
HVLLEBPAAEEAPARTQAPRYIAMPLSAHTPTLARAQAARDLSMLHNOQADPDA
HALATCTHEKRAVITGRTRELTHTLTLDAIOGTABHPRLTLTFTGOGAHPDM
GOLYTTDPHFAALDEICELORCSTONIREVMTFPDOLLDRREYTOPALFALOT
ALYRTLTAHGOAHVILGHVSEITTAHITAGVLDLPDAALITRAHMLGOLPHGGMAL
LSVQAAEHDLQALHNGVETIANVNGPTCHVLSGPTALEETLOHLEQNVHTWLVK
SHAHSALMDPMGAFRDTINTANTOPPTPLISNLITGOIADPNHLCSTPDYWDHAAH
TYRPAADVOTAHDOQRTTYLEIGHPTLTTLHHTLNDPTTLTHERPEPELTTLA
LATLHTTGHTTTPHPSHILPAQRYSLAPFORAMPMNSAHSIGSDAEARLTGIA
ROHPLETGATSIYVSGATLTGVRSLAAMPWMLADHTVAGVLEPPGAVLDLHVAE
ENGSGVETLTLHPLVLPNOGAVOLOVVEAPDERGRAATVAAAREETGEMHTH
AAVYLAGERPAAYVMDMADMAAGAMPPEGAENDVDELYAEPAAGCTDYCPATGCTG
AMRGESEVPAVRLPEVAVSGSEERFVHPALFEDALHPWVGGGPMWSDGADGCG
QSTLPLPWOGLTLYTGGAQVAVRLAPSDGAEVRAADAPACHPVLADALVLRAP
STYPLGAPAMHIDMOPVDSSTLVDRYAVVGADETAELVRAFLPDAITATA
DIALRAAPAVAPTPGPALVVALIPSPPEAPATSAPEPRAVLAHLSIVLOALE
DERLTGTTAVTAVTGVAEATOPNDVPLAGALMGLVRSQSEYPGFRRLIDVDSPE
SRAALATLASDEPOLALRAGVMLAALPTLAEAGSPAPADAPTAANGIDGYTLTGC
TGALGRVAVHLARKHGVRRMLLVSSGPNASALERELTELGHTTSCVLDLADP
ADVRAVAVPSDHPILGVVHTAGVLDGALARLTERRIDTVLAKADAVRHLHEATL
GHPILRAVLESAAGLGRPGAQSAANAANVLDAFAKLRAEBGLPVSLMDGMDERE
GMAGLDATLRLRBDGILPMPADQALDELLOALFTYRDPALVLYLWLSWALERT
AGNGASVPLRLSLVPSHRRHRTTATSVDESGPGRSALAALPAERMTVLID
TYSEQYAVGIGHASAEIDPERFHEHTGPTSLATLDEARNLGRVLCRLSTLAPHP
TPKXVAMVIAELPAAAPALGTSAEDVLTGIDOMARALALMDAEDGRKRVQRULRG
LIALDRPASGADSTRAATTVADRLDEATDEIFAFLDQLEQDPDEPLGAPDERLT
DEQP"

gene
13693..19212
/gene="nida2"
CDS
13693..19212
/gene="nida2"
/note="orf2"
/codon_start=1
/transl_table=11
/product="polyketide synthase module 3"
/protein_id="AAC6025.1"
/db_xref="GI:2558839"

/translation="MTAENDKIRSYLKRATAEILHRTKERLAELLESAREPEAVVGMAC
RYPGVATPDDELMEVLASGTDALISTFPDRDMDGLIDPDPSTPGKSYVHGGFLID
AAOFDAEPFISPREATAMPQORLLETSMELHAGIAPHTLTTRTGTGVTASQOEEGT
OSREADAKYGGHLLTGLTASVMSGRAVYTLGLOGPALVDTACSSILVALHVAQSLR
RGCDLALAGSTVMTSPVVEFESROGLAPGRCKPFADTADGTAMAGVGLVVE
OSTRGECDLALAGCTVMTSPVVEFESROGLAPGRCKPFADTADGTAMAGVGLV
LIVERLSDAERLGHVLAAYRGTAQNDGASNGTLAPGPAQOQVITDALSDAGLSAD
DIDAVEAHGCTPLGDPTEABALATTGHEROPTVWLSKSNIGSQAAGACITIKMO
KMOAHNGTLPRTLHVADPTSOVMEAGSLQILAEARPWADPRRAGISAFVSGTNA
GSHTEFCGSRSPRAEVALLOAGAGAVORLIDIPAGRPLISCESELPR
GSHTEFCGSRSPRAEVALLOAGAGAVORLIDIPAGRPLISCESELPR
TOADRODIALATETPRHKAHVITGRTRELTHTLTLDAIOGTABHPRLTLTFT
GOGAHPGMOEGLYTTDPHFAALDEICELORCSTONIREVMTFPDOLLDRREYTOPALFALOT
OPALFALOTLALYRTLTARGTOAHVLGHVSEITTAHITAGVLDLPDAALITRAHMLGOLPHGGMAL
GOLPHGGMALSVQAAEHDLQALHNGVETIANVNGPTCHVLSGPTALEETLOHLEQNVHTWLVK
SHAHSALMDPMGAFRDTINTANTOPPTPLISNLITGOIADPNHLCSTPDYWDHARHYRPAAYOT

gene
CDS

DWIDHARHVRPAADAVOTAHNGCTTYTLEIGHPTLTTLHHTLNDPTTLTHER
PEPELTTLATLHTTGHTTTLTFTSPQTHHDLPTYPQORDRYWMEPRVAQSGOPADRLRYRWMEAPEDHT
VPAERMLLGGSDPADAGILVEBOLVGHGAAYORLIDIPAGRPLISCESELPR
GSHTEFCGSRSPRAEVALLOAGAGAVORLIDIPAGRPLISCESELPR
OLMGIOVAGLELADRMGGLVDLPADAPALRGILARLVSNAPDNQVIRASGVPR
RYVGRPRYRNDMAPSGYLTITGCGALSGOVARRLALAGAPHLILAGRGSGAEP
LVDELALGAEVTVACDADADRALVSLATPEHRITLVAAGLVLDGVLGILTP
ERIDAVLAKATARHLDLTDADLDAEVLFSIVGWNGGOATYAANAALDALA
HGRARGORATSIANGWAGSGAAGAKSLARDGIDVLEPEALDLEOARSET
TLIVADMETFLSSASRRFOPLFCRIPARAVANGADSPDVPVLSLALHVAQSLA
EROLRLIDIVTEASAVLRHGSATVADEAFARAVGSDSLTVELIRRLASAGIRLS
NTLPRPPTAAAYELBOLCTESBAASHODACARTYVADERYAVVGMACRIPG
GVATPDDELVASGTDALISTFPDRDMDGLIDPDPSTPGKSYVHGGFLIDAAOF
DAEPFISPREATAMPQORLLETSMELHAGIAPHTLTTRTGTGVTASQOEEGT
OSREADAKYGGHLLTGLTASVMSGRAVYTLGLOGPALVDTACSSILVALHVAQSLR
RGCDLALAGSTVMTSPVVEFESROGLAPGRCKPFADTADGTAMAGVGLVVE
RLSDERLGHVLAAYRGTAQNDGASNGTLAPGPAQOQVITDALSDAGLSADDDA
VEAHGCTPLGDPTEABALATTGHEROPTVWLSKSNIGSQAAGACITIKMO
AMHGTPLRTLHVADPTSOVMEAGSLQILAEARPWADPRRAGISAFVSGTNA
HVLLEBPAAEEAPARTQAPRYIAMPLSAHTPTLARAQAARDLSMLHNOQADPDA
HALATCTHEKRAVITGRTRELTHTLTLDAIOGTABHPRLTLTFTGOGAHPDM
GOLYTTDPHFAALDEICELORCSTONIREVMTFPDOLLDRREYTOPALFALOT
ALYRTLTAHGOAHVILGHVSEITTAHITAGVLDLPDAALITRAHMLGOLPHGGMAL
LSVQAAEHDLQALHNGVETIANVNGPTCHVLSGPTALEETLOHLEQNVHTWLVK
SHAHSALMDPMGAFRDTINTANTOPPTPLISNLITGOIADPNHLCSTPDYWDHAAH
TYRPAADVOTAHDOQRTTYLEIGHPTLTTLHHTLNDPTTLTHERPEPELTTLA
LATLHTTGHTTTPHPSHILPAQRYSLAPFORAMPMNSAHSIGSDAEARLTGIA
ROHPLETGATSIYVSGATLTGVRSLAAMPWMLADHTVAGVLEPPGAVLDLHVAE
ENGSGVETLTLHPLVLPNOGAVOLOVVEAPDERGRAATVAAAREETGEMHTH
AAVYLAGERPAAYVMDMADMAAGAMPPEGAENDVDELYAEPAAGCTDYCPATGCTG
AMRGESEVPAVRLPEVAVSGSEERFVHPALFEDALHPWVGGGPMWSDGADGCG
QSTLPLPWOGLTLYTGGAQVAVRLAPSDGAEVRAADAPACHPVLADALVLRAP
STYPLGAPAMHIDMOPVDSSTLVDRYAVVGADETAELVRAFLPDAITATA
DIALRAAPAVAPTPGPALVVALIPSPPEAPATSAPEPRAVLAHLSIVLOALE
DERLTGTTAVTAVTGVAEATOPNDVPLAGALMGLVRSQSEYPGFRRLIDVDSPE
SRAALATLASDEPOLALRAGVMLAALPTLAEAGSPAPADAPTAANGIDGYTLTGC
TGALGRVAVHLARKHGVRRMLLVSSGPNASALERELTELGHTTSCVLDLADP
ADVRAVAVPSDHPILGVVHTAGVLDGALARLTERRIDTVLAKADAVRHLHEATL
GHPILRAVLESAAGLGRPGAQSAANAANVLDAFAKLRAEBGLPVSLMDGMDERE
GMAGLDATLRLRBDGILPMPADQALDELLOALFTYRDPALVLYLWLSWALERT
AGNGASVPLRLSLVPSHRRHRTTATSVDESGPGRSALAALPAERMTVLID
TYSEQYAVGIGHASAEIDPERFHEHTGPTSLATLDEARNLGRVLCRLSTLAPHP
TPKXVAMVIAELPAAAPALGTSAEDVLTGIDOMARALALMDAEDGRKRVQRULRG
LIALDRPASGADSTRAATTVADRLDEATDEIFAFLDQLEQDPDEPLGAPDERLT
DEQP"

Query Match 57.4%; Score 58; DB 1; Length 41097;
Best Local Similarity 74.5%; Pred. No. 0.00051;
Matches 73; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 4 CGGCTGCGAGGCTGCTGCCGACGAGCGTCAAGACTTCTCGGCCGACCGACGCG 63
D 4284 CGCCAGCGGGGACGTGCGCGCGGCGCGTCAAGGCGCTCTCGGCCGATCGCGACGA 4343
QY 64 GTGGGCTGAGCGAAGCTCGGCATGCTCTGCTCAA 101
D 4344 ACTGCTGGGCGCGGAGGCGTGGCGATGCTCTGCTCGA 4381

RESULT 13
SNA132222
LOCUS SNA132222 20394 bp DNA linear BCT 24-Apr-1999
DEFINITION Streptomyces natalensis pims1 gene.
ACCESSION AJ132222
VERSION AJ132222.1 GI:4678702
KEYWORDS pimaricin biosynthesis; pims1 gene; polyketide synthase.
SOURCE Streptomyces natalensis
ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

			AAAPAEVDPDAIAMPPOGAGLELDAACGAYERRPADFPAYGFPGLRRAAMKIGEDIVYAF VAIPREGDGNAGYLHPALFPAALHAAILLGCGGTDEAAYVPSPNNMPTLIHATGASVVRP IRPREGGTSTAIVDTACAPAVASVSILTARPIACQLOTDGRBDSLFQVDWTLHLTDER ANSALLGLKDTEGIPLFLSLQPADLDLLAATGVHDVLAELPRYGAGVESVHAATT GALLIKSMILADDREFFAASRLVFYTRGCAVSDLAGASWGVLRSALLHEHPRGELVDV DVDDADVETLVPPRALASDEPOVLYRGESEVLNIAVASOSPOTYWDSEGTVLTIGTGCG GLGSVAERHLVESEHGVSILLVSRGAEEVDNALVAECGAOVMEACDVMDAVAWA VADIABHRISAVYTACVIDDCVYESITPERLSVALPKYDAANMHERRGDLIDLP FYVFSVAGTFEGSAGQANYAGNAFLDALATHRAVGCLPAYSLAMGWMSODSGMTGTL SDADVQRIRAQMGMPITVERGLEFDLALGSAEEMALPVRLIDLALAQAQGEPOLILKG LIRPFGRRTAASAATEVEDRTAAAFAGRLTGLSAMBEREVGLVARSOIAGVIGHARETE DODPAGLDFGDSLDAVELERNRLGAVTGIRLPATLLFDPYPPALVNLHRIAPIPEETE VGPFALLGEIERMESFSGGLDIITEEMHQJAGRLEVLRANKDALRDYAAAAAHGDGSSPS DEDPDFESADSDEYFDLLDNENIGTS"			
BASE COUNT	2611 a	7486 c	7397 g	2900 t		
ORIGIN						
Query Match	57.2%	Score 57.8;	DB 6:	Length 20394;		
Best Local Similarity	76.3%;	Pred. No.	0.00066;			
Matches	71; Conservative	0;	Mismatches 22;	Indels 0; Gaps 0;		
Oy	3	C CGGCTGCAGAGCCCTGCCGACGAGCAGCGGTTCGAAGAAGCTTCGCGCCAGACCAGCG	62			
Db	5319	CCGGCGAGCGCGGGGCGGCCCCCGCGCTGCAAGCGCGTTCCGGCGGCGCGCCAGCGG	5378			
Oy	63	C GTGGCGTGCAGCACCAAGCGCTCGCCCATCTCTCT	95			
Db	5379	CACCGGCTGCAGCGAGCGGCGTTCGGCTCTCTCT	5411			
RESULT_15						
LOCUS	AB070942	27522 bp	DNA	linear BCT 11-OCT-2001		
DEFINITION	Streptomyces avermitilis polyketide-2 biosynthetic gene cluster.					
ACCESSION	AB070942					
VERSION	AB070942.1	GI:15823997				
KEYWORDS	Streptomycetes avermitilis DNA.					
ORGANISM	Streptomycetes avermitilis					
REFERENCE	1 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osone,T., Kikuchi,H., Shiba,T., Sakai,Y. and Hattori,M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites					
AUTHORS	JOURNAL	TITLE				
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)					
REFERENCE	2 21477403					
AUTHORS	Ikeda,H.					
JOURNAL	Direct Submission					
TITLE	Submitted (01-sep-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan					
E-mail:	ikedamc.pharm.kitasato-u.ac.jp,					
Fax:	+81-3-3444-6197					
FEATURES	Location/Qualifiers					
SOURCE	1. .27522					
	/organism="Streptomyces avermitilis"					
	/db_xref="taxon:33903"					
	/note="Actinobacteridae"					
	complement(1..555)					
	/codon_start=1					
	/transl_table=11					
	/product-"hypothetical protein"					
	/protein_id="BAB69213.1"					
	/db_xref="GI:15823998"					
	/translation:"MAETAAPKAADAAPRSFTLSDTTAGPVGRTGRRVRIDKOSPKA YHALVESEAIVRAVVADGALDIVALELINRVOINCAGCALSIHTFAAIRVGETTOR LGLIAMRWDETERTSRERALDALAEVTTPDADTAQGHAIYDAARAQVLTIDEISAVTW AVTIWMNRYSIMSKHPVKRPIDQ"					
CDS						

CDS
complement(584..1030)
/codon_start=1
/transl_table=11
/product="putative isomerase"
/protein_id="BAB69214.1"
/db_xref="GI:15823999"
/translation="MDEPVARSLLAELRGRGHEVYVANGALRPGDGTBMAACSAARE
VATGSDAVVCMCTGTGASTAAKKVGYRAALCTDGYATDAGRRANDANYALSLRL
TSQPLREILDAWEFAAEAGEDADDRENVAHVGRILDADRTGAGDRG"
complement(1069..1869)
/codon_start=1
/transl_table=11
/product="putative carboxylase"
/protein_id="BAB69215.1"
/db_xref="GI:15824000"
/translation="MPQLAVDAGACTYDDDEGPHDGGCVPLVFIHGWTADRRHNDHMA
AHFAKRRVVLIDLRGHEGSGSARTIDELGVDYIALDLHLEIDRFIPVGHSMGMAA
OTLAAHERLEIRLVNYSIRMTYSRGRGLMASTLPEFKLEVAANIIRAPAGGPR
REVEERYIRASSATPROVVMYVYAAMSFEVDLVGELRMPETLLVHYGYDQLPVSQM
LRMADYDAYVRYIVDAGHELPEYKRAELTSAIDLRPTGGVGYQVLG"
complement(1933..3069)
/codon_start=1
/transl_table=11
/product="probable esterase/lipase"
/protein_id="BAB69216.1"
/db_xref="GI:15824001"
/translation="WIFRTDSTTRPALRRVYRTTTLRPLHHRIDLRPAVEIIPRACD
GCRIGLTIVDSGEPDRPAVVLHGHGTSADMFLPETRNLVDALDDGYEPWLLDMWG
SGLRYNTEGOKRYDYDAALYDIPAASVHIGRIGDRPLFVAPHCISGLTSLSTMG
LVPGAGVAVSGVFEPLKLAGRTSLRMSVAGELKTRIDHILPDEFKVGMSKTYPLF
ALASKACSPPTCOOILHNSAWMGASLFEVHEHSEATHRLAELGAPALMLPHLR
RIELARSVYRNHDDHRYRELQNALDAAGRIDPVPYLINGSENGMLDQKLCHEVLR
ARKQPYLLVSTTELRIGLDTFLRGALADVFGHILDFGERR"
complement(5628..20057)
CDS

[illegible]

```
SAOSVRPDPQLRELGMDSYAVELRNKIRIGTIGAKLPATLLFDHPPTSLAEYLTTTA
LATAGPARRATAGADPARATPASDEPVAIVNSMACLPVGSVDPEGLIMLVAGROA
VGPFGRMVDVESLYDDPPEALCKSYAREGGFLDIESFPAGFFGTIPRAAMADPOQ
RLLEMAWEALERAGTYPAELAGSTTCGVYMGSDGSRLLDQLDGVCGSALSIV
ASGRSLATLGLRPAALVTDPACSSIVATITLAAOLRSGECDLALACGVLTMTYPTGF
VERSLKGLSPTRKCRNSDDADGAWAEAGMLVLRSLDARRDQVLAIVLRGTAV
NOGRSOGSLAPNGPAQEOVYIRALRELISAGAADIDVEAHGTGTIGRPIENALISE
VFQDSPRENPRLYIGLSLKNIGHAQASGVILGLKVOSLRHGTLPRTLHAGPFSQIV
WMDSGSLHLOEAVDMPSSGSEVRAGVSAGISGTNAHYIVEAPAPADPEAPAEEM
PKGRFLPSGRSEAVYNGQAAQLAOYLTEDVALPDVAHTLARHSHERETATYAARM
DELRALDALATGRTPLVPPREOTGVAFVFGHGQMPQMGVEIMWASDAFEEELAR
IGSLGELIAAAYSAGALITDAAYTVTRQAQAVVPLAKGGMLSVELPRAQVEELAV
YAGRLFAAANSAAATAVSGDAALAEVLRLEERKIPARSLTSPFASHHPMLDPE
ELLDFRADVAGARTPPPLYSTVLAEPVPGGRLLAAVNAALREVRAIOWVRMLDNG
YRYFVELSPHSIGPSTIEAANAAGIDAVVSGSIRRODQERDILRLRGLIYAAGHP
DMPVLFPTGRKVDLPYAFARERHMLAPASPGGGGLGTHVEASDPRDLRFOS
DLDLRDSRFAYLTDHRYTGDVWLGGAFLDMALAEASALODGGEVRLADVRLPRL
DEGRPAILOVLRPAEDGVDFETIASVTEGERTRWERHVSAGVTYAGIPMDARLA
VLREOCTEOVDLPGVYARLALGIDGPARGLBEGERTHSALGRLAHPAGHLH
PAVIDAFTHAALPGDAPEGAFVPAFGFGLRTPTPVWVTCELRSMAQDRTD
LRLMDEREQDLVEAEPELALSPIDGALFETTKQPRPPAPQOEPATSSNILLADETV
AVELEERLSSVPHVIVRGEAFVGBPGRYVLDPPADQOLAHLDLDAIADGPERYV
QLSALDAPALGEARTAEQARLCCSLTLHLVRLTADPREGARQLFVYVRSQAVGDS
TOYTHPOQALAMGFLAVAOEPELATLLIDLPSTDVDALMEQLRHADDERLVALRD
TGRIVPRLARTRPDDGGHGLTPPDGYLLTIRPGGLGRVYAERLVRRGARLALISG
RGADAAGWTKGLEEGVYVYHLARADYADAGITLALALAIRAGAPVAYVHAAGVLD
DATVANLTDERVULVPKVLGTTLTLELIPRGNVLFLRASAGLLSAGOSPYSAN
AFLDMAHNLSTRIDRRALSLDWGAMAGVNAVSGTADTSBGIVAFSPDSEGELF
ERVLATSRLOLAIALDWMELADADAAARRPLGLDYVPTGGASDDLVKVFAT
TDHRAVRLAEAYRVAEYCGAVDGGTTLKEGLDGLSMLVRLNAPARELGVEL
PAAAFSASDIRGLVRLVVALPERDTPPEOPADVTSEVLEVPETELRPATRDVRL
LRSQPMGAHNAVGLAVRLTPTTRETITGITRLVYRHAALRTAIYPGABGRQL
RVDRERANVPLKNTDVADYELDAGDRRLRLSPDITTAALMRFLDGGARGQTL
VFGAHHAVSDJSLLLVAGEIDAEISGTSIGDVTNNDIDLIOAQRTEGTVPGDGP
GAAAAVEMREAFHSGSRLLDITLREPARSRYASGVTVAIPDGLMERVSAASGLAV
TPAAPCLGTLTVLLAHRERERFVLAVPDTRIHADYDAVEFGVVPVPAEGAGE
PVQOVLRTRTDGLKLVAKGMFSDVPLTAKOGLVPAAPLIVEVYVRSASGRLT
NLEVLPACTGMSDLMTKTPPPAGRIRLDHNIDIDAAATTDLGEEFLRLARAYD
PTAPVPTTAASEAPAAATVLAASEVPPAALAVSEAPATEVPATVEPPAGEPAY
PPPLVAAVTAFFALNLPMLCELACNEGAGDGVYVAEPYHHVLAISLRDPSGFAD
PATAVGVTLRAADLERFPGVDALLAELTEYPAALRASETRKPLIVGPPSARA
DRLERMSSEVAEELAGLPGIAVIGPDDMRHHAVERFERTEKLAHLPTDFEOA
VALRLADVRAVRAPKVIYAVDDETMGVAEGEISPEAVDLTGPRALLARLLQWR
AAGALLAIVSNDEDTYRAYLDRPDSLAKAHPFSVLSAANGPKSRLAEARTLNTGT
DSFLTLDNNEVEIAGVRSALPEVLSVTCPTAELDEFLGRMLVPPAATAEDGLAR
FYQBERERDAVREHAGFEELQOLEVEDIRALSAADVORAEQLVRTNOFTLRARSA
DGGDLITRMRECGEVTAAARDRGDYGOIGLAVRPEGQLDVLAMMMSCRALGRVE
ERLLRMLADRAEEELGCAKVLTAERTPRANVPARRVLAALGGDDAYLLETVVTPOQL
RAFRSMQOO"
complement(20358..22082)
/codon_start=-1
/transl_table=-1
/product-"non-ribosomal peptide synthetase"
/protein_id="BAB69218.1"
/db_xref="GI:15824003"
/translacion-"MARSLVDLTKHASHQPDRTATRYLVTCGRDGEIQDITSGRLAE
KSRVAWSLDERGLASRAMLYPGLFETCYIGCLISAGVAVPGVPQGRSQNHRA
LTRMKRLIADADAKVILIGREVIIAALGTMAEHLPELAETICVATEDIPDAAGSWREP
DLTADSVAFLOQYSGTSAPRGVMTNGNLDNBERVITERMGTPTDIYADHLEFVS
WLPVYHDMGLIGPVLTNTVYLGATATLESPLHFOOPORMLTALSHYRHPSTSGPNFAY
ELICKOATPDLLDGLDISRWKVAFNGAEPAAADAANVSGSPGCVTLVINADPERQOGL
AEATLAWTGGSVAPPTLIEAPETGPHAGAADAANVSGSPGCVTLVINADPERQOGL
PQCEVGETIYVLSAAGVYKRNLTATRESRATLKNREGFLRTGDLGFLRDELFT
GRUKDLIVIDGRNNHYPODELISAEHMPALRPGCTAIFSVDGVEGEQPIVAEAAPE
AADEFKRTDIRISAIGEAHGLSVRDVIVHPGTIKTSSGKIQRRASTRAYLDGTLS
VVGAAATR"
22404..23084
/codon_start=-1
/transl_table=-1
/product-"putative methyltransferase"
/protein_id="BAB69219.1"
/db_xref="GI:15824004"
```


[illegible]

RESULT 7
AQ797032

LOCUS		437 bp	DNA linear	GSS 04-AUG-1995
DEFINITION	AB079703.2			
ACCSSION	AB079703.2	CUGI Rice BAC Library	Oryza sativa genomic clone	
FEATURES	nbjbb0071n03r,	DNA sequence.		
ORIGIN	nbjbb0071n03r,	DNA sequence.		

ACCESSION	AQ/97032
VERSION	AQ97032.1
KEYWORDS	GSS.
SOURCE	Ornith. sativa

SOURCE
ORGANISM

REFERENCE
1 (bases 1 to 437)

AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA

Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: GGAAACACCTATGACCATG
 Class: BAC ends
 High quality sequence stop: 233.
 Location/Qualifiers
 1..437

High quality sequence stop: 233.

FEATURES

Source

```

54 a      148 c      149 g      85 t      1 Others
BASE COUNT
ORIGIN
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="ndxb0071N03r"
/clone_ldb="CUGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notice=Vector: pbeloBAC11, Site_1: HindIII, Site_2:
HindIII: Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

```

	Query Match	33.1%	Score 33.4	DB 17	length 437	
	Best Local Similarity	65.3%	Pred. NO. 71			
	Matches 49	Conservative 0	Mismatches 26	Indels 0	Gaps 0	
0y	15 CCGGGGCTTTCATGGGGCCGCCGACAGCCCGTGCCGTGCATGGCCGCTTCGCA	74				

Db	Qy
278	75
CGTGGCGCTCCGCTCCGAGGGGTTGCAGAGCTCCGTCGCCCTCAGACCCGCGCGTTC	CTTGGGCTCGACTC
337	89
CGTGGCGCTCCGCTCCGAGGGGTTGCAGAGCTCCGTCGCCCTCAGACCCGCGCGTTC	CTTGGGCTCGACTC
338	352
CGTGGCGCTCCGCTCCGAGGGGTTGCAGAGCTCCGTCGCCCTCAGACCCGCGCGTTC	CTTGGGCTCGACTC

RESULT 8

LOCUS	AV6288327	384 bp	mRNA	EST 15-DEC-2000
DEFINITION	AV6288327 Chlamydomonas reinhardtii	5% to 0.04%	CO2	Chlamydomonas
	reinhardtii cDNA clone LC0477f08_r	5'		mRNA sequence.
ACCESSION	AU628832			

ACCESSION	AV628827
VERSION	AV628827

KEYWORDS
COUNCIL

SOURCE
ORGANISM
Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales

REFERENCES

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 504)	Asamiya, E., Mura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S.	Generation of expressed sequence tags from low-CO ₂ and high-CO ₂ adapted cells of <i>Chlamydomonas reinhardtii</i>

JOURNAL
DNA Res. 7 (5), 305-307 (2000)
MEDLINE
20539644

COMMENT

Kazusa DNA Research Institute
The First Laboratory for Plant Gene Research
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asami@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>

FEATURES
SOU

```

/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCU047f08_r"
/clone_id="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI; The CDNA library was constructed from cells cultured
in a carbon stress acclimated condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

```

Query Match	32.9%	Score 33.2	DB 10	Length 384
Best Local Similarity	64.1%	Pred. No. 78		
Matches 50; Conservative	0	Mismatches 28	Indels 0	Gaps 0

Qy 2 TGCAGCCGAGATCGCCCGCTTTCATGGGGCGCCGAGCGCCGTGCCCGTCATC 61
|||||
Db 262 TGCAGCCAGGTTGCACATGCACGTGGCTTGGCCGAGCGCCGAGCATCAAGC 321

QY	62	GGCCGCTGTCGGACTGG	79
Db	322	AGCTGAGTCCGAGCTGG	339

RESULT 9
AVA32901
00000001
000 1-
-0000 1-0000
PCN 22-AUG-2000

LOCUS
DEFINITION

PM021e05_r 5', mRNA sequence.

ACCESSION	AV432901
VERSION	AV432901.1
	GI:8588126

KEYWORDS

SOURCE ORGANISM

Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;

REFERENCE

AUTHORS Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and

TITLE	Tabata, S.
JOURNAL	Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, <i>Porphyra yezoensis</i>
MEDLINE	DNA Res. 7, 223-227 (2000)
COMMENT	20363100 Contact: Erika Asamizu

Kazusa DNA Research Institute
The First Laboratory for Plant Gene Research
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>

FEATURES	
source	Location/Qualifiers
1..502	
/organism="Porphyra yezoensis"	
/strain="TU-1"	
/db_xref="taxon:2788"	
/clone="PM021e05.r"	
/clone_1lb="Porphyra yezoensis_TU-1"	
/note="Vector: pbluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"	
37 a	206 c 186 g 72 t 1 others

Query Match	32.9%	Score	33.2	DB	10	length	502
Best Local Similarity	59.6%	Pred. No.	83				
Matches	56	Conservative	0	Mismatches	38	Indels	0
						Gaps	0

9y 7 GCCGAGATCGCGGGCGCTTTCATGTGGGCGCGCGAGCGCCGCGCCCTCGATCGGGCG 66
 |||||
 80 GCCCGCCCGCTGCGCCCACTGATGGGGCGCGCGCGATGTGGCGCGCGCTCCACGGGCTG 139

0y	CTCTCGGACTTGGGGCCCTCGACTCGCTCAAGGCGG	100
67	CTTCTGGGCTTGGGGCCCTCGACTCGCTCAAGGCGG	100
140	CGGGTGGCCCTCGACCCGCGACGCGATCGAGGCGG	173

RESULT 10	LOCUS	BO652574	955 bp	mRNA	linear	EST 15-JUL-2002
AGENCOURT	8511766	NIH_MGC_100	Homo sapiens	CDNA clone	IMAGE:629685	
5', mRNA	sequence.					

VERSION	B0652574.1	GI:21776746
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 955)	
TITLE	NIH-MGC http://mgc.ncl.nih.gov/ .	
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	

Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rudin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: Agencourt Bioscience Corporation
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LRCM2505 row: f column: 22
 High quality sequence start: 9
 High quality sequence stop: 430.
 Location/Qualifiers
 1..955
 FEATURES
 source

/db_xref="taxon:9606" /taxon="Homo sapiens"
 /clone="IMAGE:6296853" /clone="1b"
 /tissue_type="NH_KGC_100" /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)" /lab_host="DH10B (phage-resistant)"
 note="Organ: liver; Vector: pORF5; Site.1: XhoI; Site.2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor: GGCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library.

Query Match	32.9%;	Score	33.2;	DB	14;	Length	955;
Best Local Similarity	59.6%;	Pred. No.	97;				
Matches	56;	Conservative	0;	Mismatches	38;	Indels	0;
						Gaps	0;

8 CCGAGATGCGCGCGCTTTCATGCGGCGCGCGAGACGCGGTGCCGTGATCGCGCG 67
 853 CCGGACGCTGCGCGCTTCCGGCTGGCGGTACGACTGCCGGCGCGCGTCCGCGCGCGGT 794

Qy	68	733	760
Db	TCGCGACTTGCGCCTCGACCTCGCTACGGCGGT	GGCGAGTCACCTCTGCGCTGTGCGGCACGGCGGT	

RESULT 11				
CNS01087/c				
LOCUS	CNS01087	961 bp	DNA	linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC			
	BAC033007 of Drosophila library from Drosophila melanogaster (fruit			

ACCESSION	fly), genomic survey sequence.
VERSION	AL098641
KEYWORDS	AL098641.1 GI:5610252
SOURCE	GSS.
ORGANISM	Drosophila melanogaster. Drosophila melanogaster.

REFERENCE
AUTHORS
TITLE
JOURNAL

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 961)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL

FEATURES	COMMENT
Location/Qualifiers	<p>Web : www.genoscope.cns.fr</p> <p>Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pMelBAC11.</p>

source	1. : 301			
	/organism="Drosophila melanogaster"			
	/db_xref="taxon:7227"			
	/clone="BACN03H07"			
	/clone_1bp="DrosBAC"			
	/plasmid="pbeloBAC11"			
	/note="end : Sp6"			
BASE COUNT	180 a	208 c	163 g	174 t
ORIGIN	236 others			

Query Match	32.9%	Score 33.2	DB 17	length 961
Best Local Similarity	26.5%	Pred. 0.97		
Matches	26	Conservative	44	Mismatches
			28	Indels
				Gaps
				0

Oy 1 GTGACAGCCCGAATCCGCAGCGGTGCTTTCAATGGGCGGCCGCGAGCCGTCGCCCTCAGAT 60
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 950 GGGSSSSSGSSSCCSCSBSBSSSSCCGCGGBSBBCBSSCCSCCSCTmTSCSSSGSSC 891

Oy	61	CGGCGCTGTGCACACTTTGGGCGTTCGAATCGGTCAACGCC	98
		: :: : : : : : :	
Db	890	CGSGCGCCCTYSSGKSCGSGSGKSTBYBEBDCSCSSS	853

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	JOURNAL	COMMENT
AV924405	AV924405	542 bp	mrna	linear	EST 18-JAN-2002					
LOCUS	AV924405 K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA clone basd14g21 5', mRNA sequence.									
DEFINITION	AV924405									
ACCESSION	AV924405									
VERSION	AV924405.1	GI:18220184								
KEYWORDS	EST.									
SOURCE	Hordeum vulgare subsp. vulgare.									
ORGANISM	Hordeum vulgare subsp. vulgare									
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae									
AUTHORS	1 (bases 1 to 542)									
JOURNAL	Sato, K., Salscho, D. and Takeda, K.									
COMMENT	Bailey EST sequencing project in NIG and Okayama Univ unpublished (2002)									
FEATURES	Contact: Tadasu Shih-1									
source	National Institute of Genetics									
	111 yata, Mishima, Shizuoka 411-8540, Japan									
	Tel: 81-559-81-6856									
	Fax: 81-559-81-6855									
	Email: tshini@genes.nig.ac.jp.									
	Location/Qualifiers									
	1..542									
	/organism="Hordeum vulgare subsp. vulgare"									
	/cultivar="Haruna Nijo"									
	/db_xref="taxon:112509"									
	/clone="basd14g21"									
	/clone_1kb="K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves"									
	/tissue_type="seedling leaves"									
	/dev_stage="second leaf stage"									
BASE COUNT	90 a 200 c 169 g 83 t									
ORIGIN										
Query Match	32.7%; Score 33; DB 10; Length 542;									
Best Local Similarity	58.8%; Pred. No. 95;									
Matches	57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;									
1	GTGACGCGCGAGATCGCGCGCTTTTCATGAGGCGCGCGAGCGCGCGCTCGCAT 60									
111	111 111 111 111 111 111 111 111 111 111									
404	GTGTCGGCGGACGCGCGCGCGCTGTATCTGACGCGCGACCGGTGAGCGCTCGAG 463									
61	CGCGCGCTGTGACCTGTGAGCGCTGCATCGCGTCAAGG 97									
1111111	1111111 1111111 111 111 111									
464	CGCGCGCTTCTTGAGATGTGTCCGCTCAGACGACCG 500									
RESULT 13										
LOCUS	AL818682	513 bp	mrna	linear	EST 15-JUL-2002					
DEFINITION	AL818682 1:125 Triticum aestivum cDNA clone B06_1125_plate_12, mRNA sequence.									
ACCESSION	AL818682									
VERSION	AL818682.1	GI:21830282								
KEYWORDS	EST.									
SOURCE	bread wheat.									
ORGANISM	Triticum aestivum									
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae									
AUTHORS	1 (bases 1 to 513)									
COMMENT	Wilson, I., Beswick, R., Shepherd, S., Barker, G., Parker									

FEATURES	Source	1. .513 /organism="Triticum aestivum" /cultivar="mercia" /db_xref="taxon:4565" /clone="B06_1125_plate_12" /clone_1lb="1:125" /tissue_type="embryo" /dev_stage="30 days post anthesis"	103 a 135 c 172 g 102 t 1 others
BASE COUNT	ORIGIN		
Query Match	Best local Similarity	32.5%, Score 32.8; DB 9; Length 513; 59.8%; Pred. No. 1e+02;	
Matches	55; Conservative	0; Mismatches 37; Indels 0; Gaps 0;	
QY	1	GGCAGCGCCGAGATCGCCCGCTGCTTCATCAGGGCGCCGCGAGCGCCGTCGCAT 60	
DB	185	GTGGGGGTGCGGGGGCCGCGCGCGCTGGCTTCAGGGAGCGCGCGAGGCCCGCTGGGGGTACA 244	
QY	61	CGGCGCGTGTGCGAGACTTGAGGCCCTGCATCGCT 92	
DB	245	ACGGCGCGCGCGCGCTGGCAGCGACCAACGCT 276	
RESULT 14	BM909387	1565 bp	mrna
LOCUS	AGENCOURT_6640954	NIH_MGC_99	Homo sapiens cdna clone IMAGE:5433958
DEFINITION	5', mRNA sequence.		
ACCESSION	BM909387		
VERSION	BM909387.1	GI:19359766	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1565)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgsbbs@email.nih.gov Tissue Procurement: Lou Straut cdna library Preparation: Rubin Laboratory DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.lnl.gov Plate: LICM1910 row: d column: 23 High quality sequence start: 32 High quality sequence stop: 137. Location/Qualifiers 1. .1565 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5433958" /clone_1lb="NIH_MGC_99" /tissue_type="lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cdna made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAAG(g). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
BASE COUNT	73 a 73 c 376 g 401 t 19 others		

APPLICANT: Sherman, D.H.

```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match 38.2%; Score 38.6; DB 10; Length 36778;
Best Local Similarity 61.4%; Pred. No. 0.004;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATTGAGGCGCGCGAGCGCCGCGTCGAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34632 GTGATGCGGCGAGCGCGCTTCGTCGTCGCGACTCGCCCGAAGAGTCCCGTCGAC 34691

QY 61 CGGCGCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
   || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34692 CGGCGCGCTGTGCGAGATCGGCTTCGACTCGCTCAGCGCGGT 34732
```

```
RESULT 3
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32
```

```
Query Match 33.5%; Score 33.8; DB 10; Length 11220;
Best Local Similarity 58.4%; Pred. No. 0.097;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATTGAGGCGCGCGAGCGCCGCGTCGAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10726 GTCCGTAACGACCTGGCGACCTCTGGGACAGGACCCCGGCGGTGGACTGGAG 10785

QY 61 CGGCGCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10786 CGGCGCTTCGCGACACCGGTTGCACTCGCTCAGCGCGGT 10826
```

```
RESULT 4
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
```

```
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30
```

```
Query Match 33.5%; Score 33.8; DB 10; Length 13842;
Best Local Similarity 58.4%; Pred. No. 0.096;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATTGAGGCGCGCGAGCGCCGCGTCGAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7501 GTACGGCGGAGCGCGCGCTTCGTCGTCGCGACTCGCCGAGAGACGTCCCGCGAC 7560

QY 61 CGGCGCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
   || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7561 CGGCGCTTCAAGAGATCGGCTTCGACTCGCTCGCGCGGT 7601
```

```
RESULT 5
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1
```

```
Query Match 30.3%; Score 30.6; DB 10; Length 15872;
Best Local Similarity 56.4%; Pred. No. 0.79;
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGGCGGAGATCGGCGCGTGTTCATTGAGGCGCGCGAGCGCCGCGTCGAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2822 GTGCGGCGACAGCTCGCGCGCTCTCGACTGACAGACCCGCGGTGCACTGGCG 2881

QY 61 CGGCGCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2882 CTCACCTTCAAGAGATCGGCGCTTCGACTCCCTCATGTGCGGT 2922
```

```
RESULT 6
US-09-815-242-4056
; Sequence 4056, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4056
; LENGTH: 3189
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4056
```

```
Query Match          30.1%; Score 30.4; DB 10; Length 3189;
Best Local Similarity 57.3%; Pred. No. 0.92; Mismatches 0; Gaps 0;
Matches 55; Conservative 0; Indels 0;
```

```
QY 1 GTGCAGGCGCAGATCGCGCGCTTTTCATGGGGCGCGGAGCGCGCTGCCGTCGAT 60
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1063 GTGTGTGTGTGATCCTTCCTTCGACAGACCTGGCGCGCTGCATCCGCTGGCGCG 1122
QY 61 CGGCGCGTGTGCGACTTGGCGCTCGACTCGCTACG 96
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1123 GTGCGGCTGTGCTGATCGGACACTTCGCGGTGATG 1158
```

```
RESULT 7
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474
```

```
Query Match          29.3%; Score 29.6; DB 10; Length 3209;
Best Local Similarity 56.0%; Pred. No. 1.6;
```

```
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 2 TGCAGGCGGAGATCGCGCGTGTTCATGGGGCGCGGAGCGCGCTGCCGTCGATC 61
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1215 TGTGTGAGGCCGTGGCACATCTTGGGCATCCGCGACTTGGCTGTCAACTGAGACA 1274
QY 62 GCGCGCTGTGCGACTTGGCGCTCGACTCGCTACGCGCGT 101
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1275 GCTCAGTGGCGGACCTGGCGCTGAGACTCGCTCATGAGCGT 1314
```

```
RESULT 8
US-09-815-242-1227/c
; Sequence 1227, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseldeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1227
; LENGTH: 99
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-1227
```

```
Query Match          29.1%; Score 29.4; DB 10; Length 99;
Best Local Similarity 56.8%; Pred. No. 1.9;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 2 TGCAGGCGGAGATCGCGCGTGTTCATGGGGCGCGGAGCGCGCTGCCGTCGATC 61
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 99 TGTGTGTGTGATCCTTCCTTCGACAGACTGGCGCGCTGATATCCCGTGGCGCGG 40
QY 62 GCGCGCTGTGCGACTTGGCGCTCGACTCGCTACG 96
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 39 TGGCGGTGTGCTGATCGGACACTTCGCGGTGATG 5
```

```
RESULT 9
US-09-815-242-7980
; Sequence 7980, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseldeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```

1  APPLICANT: Trawick, John D.
2  APPLICANT: Carr, Grant J.
3  APPLICANT: Yamamoto, Robert T.
4  APPLICANT: Xu, H. Howard
5  TITLE OF INVENTION: Identification of Essential Genes In
6  TITLE OF INVENTION: Prokaryotes
7  FILE REFERENCE: ELITRA.011A
8  CURRENT APPLICATION NUMBER: US/09/815,242
9  CURRENT FILING DATE: 2001-03-21
10 PRIOR APPLICATION NUMBER: 60/191,078
11 PRIOR FILING DATE: 2000-03-21
12 PRIOR APPLICATION NUMBER: 60/206,848
13 PRIOR FILING DATE: 2000-05-23
14 PRIOR APPLICATION NUMBER: 60/207,727
15 PRIOR FILING DATE: 2000-05-26
16 PRIOR APPLICATION NUMBER: 60/242,578
17 PRIOR FILING DATE: 2000-10-23
18 PRIOR APPLICATION NUMBER: 60/253,625
19 PRIOR FILING DATE: 2000-11-27
20 PRIOR APPLICATION NUMBER: 60/257,931
21 PRIOR FILING DATE: 2000-12-22
22 PRIOR APPLICATION NUMBER: 60/269,308
23 PRIOR FILING DATE: 2001-02-16
24 NUMBER OF SEQ ID NOS: 14110
25 SOFTWARE: FastSeq for Windows Version 4.0
26 SEQ ID NO 7960
27 LENGTH: 1413
28 TYPE: DNA
29 ORGANISM: Pseudomonas aeruginosa
30 FEATURE:
31 NAME/KEY: CDS
32 LOCATION: (1)..(1413)
33 US-09-815-242-7980

```

```

Query Match      29.1%; Score 29.4; DB 10; Length 1413;
Best Local Similarity 58.6%; Pred. No. 1.8;
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0

OY      15  GCGCGCGCTGCTTCATCGGGCGCGCGCGCGCTGCCCGTCGATCGGCCGCTGCGGA 74
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       969  CCGCGCGCGCGCTGTCGAGATCGACGCGCGCGCGCGCTGCCCGCTCTACGCGTTGCTGATCAC 1028
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY       75  CTTGGGCGCTCGACATCGCTCAGCGCGGT 101
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1029  CCTGTGCTTCGCGCCCTGCTCTGTTGCT 1055
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-09-919-172-93
; Sequence 93, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919.172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 93
; LENGTH: 1274
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 002455.1
US-09-919-172-93

Query Match      28.7%; Score 29; DB 10; Length 1274;
Best Local Similarity 55.4%; Pred. No. 2.4;

```

Matches	56;	Conservative	0;	Mismatches	45;	Indels	0;	Gaps	0;
Oy	1	GGCAAGCCGATATCGCGCGCTGCTTTCATATGGGGCGCCGAGGCGCCCTGGCCGCGAT	60						
Db	278	GGCGCGAGAGTACGCTTGGCGCACCTGGAGAAAGGGCGCGAGCGCCGTGAGTTCTTGGAG	337						
Oy	61	CGGCGCGTGTGAGACTTGGGCGCTCGACTGTGCTACAGGCGGT	101						
Db	338	CTGCTGGGGCGCGACTACGGCGCCCGGACGAGAGGCGCAGT	378						

```

RESULT 11
US-09-861-289-15
; Sequence 15, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-15

```

```

Query Match          27.7%; Score 28; DB 10; Length 1140;
Best Local Similarity 58.3%; Pred. No. 4.5;
Matches    49; Conservative    0; Mismatches   35; Indels     0; Gaps      0;

QY       9 CGAGATCGCCGCGCTGCATTTCATGGGGCCGCCGACGCGGTGCCCTGCATCGGCCGCT 68
           || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       282 CCGCAGCTGGGCTCGGGGTGTCCGCCACCAGCGGCGCACCCCGTGGCCCGTAGAGCCGACGA 341
           || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY       69 CTCGACTTTGGGCGCTTCGACTGCT 92
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       342 GGACCACCCCACCTCGACCCGCT 365

RESULT 12
US-09-749-728B-12/c
; Sequence 12, Application US/09749728B
; Patent No. US20020142457A1
; GENERAL INFORMATION:
; APPLICANT: Umezawa, Akihiro
; APPLICANT: Hata, Jun-ichi
; APPLICANT: Fukuda, Keiichi
; APPLICANT: Ogawa, Satoshi
; APPLICANT: Sakurada, Kazuhiro
; APPLICANT: Gojo, Satoshi
; APPLICANT: Yamada, Yoji
; TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIO
; FILE REFERENCE: 00766, 000043
; CURRENT APPLICATION NUMBER: US/09/749, 728B
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: H11-372826
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: PCT-JP00-01148
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT-JP00-07741
; PRIOR FILING DATE: 2000-11-02
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver.2.0
; SEQ ID NO 12
; LENGTH: 1326

```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: (1)..(1329)
US-09-749-728B-12
```

```
Query Match
Best Local Similarity 55.0%; Pred. No. 4.5; Length 1326;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 1 GTGAGCGCCGAGATCGCGCGCTTTTCATGAGGCGCCGCGAGCGCCGTCGTCGAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 470 GAGCGCCGAGAGCGCGCGCCGCTACTGTCGCGCGCCGCGACGCGCGCGCGCT 411
QY 61 CGGCGCGTGTGCGACTTGGGCGCTGCGACTCGCTCACGGCGG 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 CGCGCGCCACTGTGTAGCGCGACGCTTCCGCGCGCGCGG 371
```

RESULT 13

```
US-09-861-289-3/C
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3
```

```
Query Match
Best Local Similarity 27.7%; Score 28; DB 10; Length 13613;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
QY 9 CGAGATCGCGCGCGCTTTCATGAGGCGCGCGAGCGCGCGCTGATCGCGCGCT 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7688 CGCAGCTGTGCTGCGGTGTGCGCAGCGCGCGACCCCGCTCGACGCGCACA 7629
QY 69 GTGCGACTTGGCGCTGCGACTGCT 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7628 GGACCAACCCACCTTGACCGCGCT 7605
```

```
RESULT 14
US-09-967-477B-3
; Sequence 3, Application US/09967477B
; Patent No. US20020156254A1
; GENERAL INFORMATION:
; APPLICANT: Xiao Qiu
; APPLICANT: Haiping Hong
; TITLE OF INVENTION: FAD4, FAD5, FAD6, AND FAD6, NOVEL
; FILE REFERENCE: BNZ-001
; CURRENT APPLICATION NUMBER: US/09/967,477B
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/236,303
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/297,562
; PRIOR FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Thraustochytrium sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1320)
US-09-967-477B-3
```

```
Query Match
Best Local Similarity 27.3%; Score 27.6; DB 9; Length 1320;
Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 12 GATCGCGCGCGCTTTCATGAGGCGCGCGAGCGCGCGTCCGTCGATGCGCGCTGTC 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 GATCCGCGACATGATTTTACGCGCGTGTGAGATGTGCGCGCTTTCGCGCTGTCG 425
QY 72 GGACTTGGGCGCTGCGACTGCTACAGCGCGT 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GCTCATGTCCAGGCGCTGCGCCACCTCGCT 455
```

RESULT 15

```
US-09-960-352-8797
; Sequence 8797, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagapan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8797
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 38-LIB2809-003-Q1-E1-B6
US-09-960-352-8797
```

```
Query Match
Best Local Similarity 27.1%; Score 27.4; DB 10; Length 446;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 2 TGCAGCGCGAGATCGCGCGCTTTCATGAGGCGCGCGAGCGCGCGTCCGTCGATC 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 TGCAGAGTGTTCCTCACACGCTGTGCGCGCGCGCGTGTGCGCGCGCTGCG 135
QY 62 GCGCGCTGTGCGACTTGGGCGCTGCGACTGCTCA 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 AGCTGACAGTGCATCTGTGCTGCACTCCGCGCA 168
```

```
Search completed: November 6, 2002, 20:31:09
Job time : 46.125 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_31170_31270
Perfect score: 101
Sequence: 1 gtagcagccgcagatcgcgcg.....ctgcactcgctcagcgcg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	71989	4	US-09-443-501A-2
2	99.4	98.4	68750	3	US-09-335-409-1
3	99.4	98.4	68750	4	US-09-568-102-1
4	99.4	98.4	68750	4	US-09-567-969-1
5	99.4	98.4	68750	4	US-09-568-480-1
6	99.4	98.4	68750	4	US-09-568-486-1
7	99.4	98.4	68750	4	US-09-568-472-1
8	99.4	98.4	68750	4	US-09-567-899-1
9	40.4	40.0	80161	3	US-09-036-987A-1
10	40.4	40.0	80161	4	US-09-370-700-1
11	40.2	39.8	561	3	US-09-154-083-13
12	38.6	38.2	4041	4	US-09-105-537-36
13	38.6	38.2	36778	4	US-09-105-537-5
14	38.6	38.2	38506	3	US-09-320-878-19
15	38.6	38.2	50937	4	US-09-428-517-1
16	37	36.6	43280	2	US-08-804-227C-1
17	37	36.6	44377	2	US-08-804-227C-7
18	37	36.6	44377	2	US-08-804-198-1
19	35.4	35.0	1434	4	US-09-434-288-3
20	35.4	35.0	4403765	4	US-09-103-840A-2
21	33.8	33.5	11220	4	US-09-105-537-32
22	33.8	33.5	13842	4	US-09-105-537-30
23	33.8	33.5	33529	4	US-09-144-085-3
24	32.2	31.9	11219	1	US-07-642-734C-1
25	32.2	31.9	11219	1	US-08-439-009A-1
26	31.2	30.9	23673	4	US-09-773-816-1
27	31.2	30.9	441529	4	US-09-103-840A-1

28	30.6	30.3	1681	4	US-09-434-288-7	Sequence 7, Appl
29	30.6	30.3	15872	4	US-09-105-537-1	Sequence 1, Appl
30	29.6	29.3	8460	4	US-08-469-005A-9	Sequence 9, Appl
31	29.6	29.3	8519	4	US-09-261-807-1	Sequence 1, Appl
32	29.6	29.3	20235	1	US-07-642-734C-3	Sequence 3, Appl
33	29.6	29.3	20235	3	US-08-439-009A-3	Sequence 3, Appl
34	28	27.7	1140	4	US-09-105-537-15	Sequence 15, Appl
35	28	27.7	1476	4	US-09-434-288-12	Sequence 12, Appl
36	28	27.7	3292	3	US-09-320-878-22	Sequence 22, Appl
37	28	27.7	13613	4	US-09-105-537-3	Sequence 3, Appl
38	28	27.7	28958	1	US-08-258-261B-6	Sequence 6, Appl
39	28	27.7	28958	1	US-08-456-837-6	Sequence 6, Appl
40	28	27.7	28958	1	US-08-457-342-6	Sequence 6, Appl
41	28	27.7	28958	1	US-08-457-342-6	Sequence 6, Appl
42	28	27.7	28958	1	US-08-458-076A-6	Sequence 6, Appl
43	28	27.7	28958	1	US-08-764-233A-4	Sequence 4, Appl
44	28	27.7	28958	1	US-08-457-335A-6	Sequence 6, Appl
45	28	27.7	28958	1	US-08-729-214-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Zierman, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
; US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5,1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGGCGCGAGATCGCGCGCTTCATCGGCGCGCGAGCGCGCGCGCGAT 60
DB 31170 GTGAGGCGCGAGATCGCGCGCTTCATCGGCGCGCGAGCGCGCGCGCGAT 31229
QY 61 CGGCGCGTGTGCGACTGTGGCGCTGCTCGCTCGCTCGCTCGCTCGCTCGCT 101
DB 31230 CGGCGCGTGTGCGACTGTGGCGCTGCTCGCTCGCTCGCTCGCTCGCTCGCT 31270

RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
```

```
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
```

```
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 60
|||||
```

```
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
```

```
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 60
|||||
DB 36782 GTGCAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCCGAGCGCGCTGCCGTGCAT 36841
```

```
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
```

```
DB 36842 CGGCCGCTGTGGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCAGGCCAGATCGCGCGCTGCTTTCATGAGCGCCGAGCGCCGTCGCAT 60
Db 36782 GTGCAGGCCAGATCGCGCGCTGCTTTCATGAGCGCCGAGCGCCGTCGCAT 36841
QY 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 36882

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCAGGCCAGATCGCGCGCTGCTTTCATGAGCGCCGAGCGCCGTCGCAT 60
Db 36782 GTGCAGGCCAGATCGCGCGCTGCTTTCATGAGCGCCGAGCGCCGTCGCAT 36841
QY 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
```

```
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 36882

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelbach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIORITY FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.4e-18;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGCAGGCCAGATCGCGCGCTGCTTTCATGAGCGCCGAGCGCCGTCGCAT 60
Db 36782 GTGCAGGCCAGATCGCGCGCTGCTTTCATGAGCGCCGAGCGCCGTCGCAT 36841
QY 61 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 101
Db 36842 CGGCCGCTGTGCGACTTGGGCGCTCGACTCGCTCACGGCGGT 36882

RESULT 9
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patli J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28, 479
; REFERENCE/DOCKET NUMBER: 50, 608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

Query Match          40.0%; Score 40.4; DB 3; Length 80161;
Best Local Similarity 63.3%; Pred. No. 0.0069;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGACAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGCGCGCGCGCGAT 60
    ||| ||||| ||||| ||| ||||| ||| ||||| ||| |||||
DB 75463 GTCCGCGCGGATATCGCGTGTACTGGGCGACGCGAGTTCAGAGCATGTGACATCGAG 75522
    ||| ||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGCTCAGCGC 98
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 75523 AAGCCTTGGCGAGCTGGGTTTCGACTCGTGCAGCGC 75560
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||

RESULT 10
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madhuri, Krishnamurthy
; APPLICANT: Treadway, Patli J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-1

Query Match          40.0%; Score 40.4; DB 4; Length 80161;
Best Local Similarity 63.3%; Pred. No. 0.0069;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGACAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGCGCGCGCGCGAT 60
    ||| ||||| ||||| ||| ||||| ||| ||||| ||| |||||
DB 75463 GTCCGCGCGGATATCGCGTGTACTGGGCGACGCGAGTTCAGAGCATGTGACATCGAG 75522
    ||| ||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGCTCAGCGC 98
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 75523 AAGCCTTGGCGAGCTGGGTTTCGACTCGTGCAGCGC 75560
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||

RESULT 11
US-09-154-083-13/C
; Sequence 13, Application US/09154083
; Patent No. 6150513
```

```

; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 13
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; US-09-154-083-13

Query Match          39.8%; Score 40.2; DB 3; Length 561;
Best Local Similarity 62.4%; Pred. No. 0.008;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 GTGACAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGCGCGCGCGTGCAT 60
    ||| ||||| ||||| ||| ||||| ||| ||||| ||| |||||
DB 371 GTGGGCGACAGAGGCGCGCATCGCGTCGCGCAAGGCTCCGAGCGCGCGTGGACGCCGAC 312
    ||| ||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGCT 101
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 311 AAGCGCTTCGCGAGATCGGCTTCGACTCGTGCAGCGCGCGCT 271
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||

RESULT 12
US-09-105-537-36
; Sequence 36, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-36

Query Match          38.2%; Score 38.6; DB 4; Length 4041;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 GTGACAGCCGAGATCGCGCGCTGCTTTCATGAGGCGCGCGCGCGCGCGCGTGCAT 60
    ||| ||||| ||||| ||| ||||| ||| ||||| ||| |||||
DB 2851 GTGATGCGGCGAGCGCGCTCCGCTCCGCTCGACTCGCCCGGAGAGGTCCCGCTGCAC 2910
    ||| ||| ||| ||| ||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGCGCGTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGCT 101
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| |||||
DB 2911 CGCCCGCTGCGGAGATCGGCTTCGACTCGCTGACCGCGCT 2951
    ||| ||| ||| ||| ||| ||||| ||||| ||||| |||||

RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
```

```
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5
```

```
Query Match          38.2%; Score 38.6; DB 4; Length 36778;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGAGGCCGAGATCGCGCGCTGCTTCATGAGGGCCCGCGAGCCGCCGTGCAT 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 34632 GTGATGGCGAGCGCGCTCCGCTCCGGTCCGACTCCGCGGAGAGAGTCCCGTCGAC 34691
    || ||||| ||| | | | | | | | | | | | | | | | | | | | | |
```

```
QY 61 CGGCCGCTGTGCGACTGGGCGCTCGACTCGCTACGCGCGT 101
    || ||||| ||| | | | | | | | | | | | | | | | | | | | | |
Db 34692 CGCCCGCTGGCGGAGATCGGCTTCGACTCGCTGACCGCCGT 34732
```

```
RESULT 14
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19
```

```
Query Match          38.2%; Score 38.6; DB 3; Length 38506;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGAGGCCGAGATCGCGCGCTGCTTCATGAGGGCCCGCGAGCCGCCGTGCAT 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 32774 GTGATGGCGAGCGCGCTCCGCTCCGGTCCGACTCCGCGGAGAGAGTCCCGTCGAC 32833
    || ||||| ||| | | | | | | | | | | | | | | | | | | | | |
```

```
QY 61 CGGCCGCTGTGCGACTGGGCGCTCGACTCGCTACGCGCGT 101
    || ||||| ||| | | | | | | | | | | | | | | | | | | | | |
Db 32834 CGCCCGCTGGCGGAGATCGGCTTCGACTCGCTGACCGCCGT 32874
```

```
RESULT 15
US-09-428-517-1
; Sequence 1, Application US/09428517
```

```
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1
```

```
Query Match          38.2%; Score 38.6; DB 4; Length 50937;
Best Local Similarity 61.4%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 1 GTGAGGCCGAGATCGCGCGCTGCTTCATGAGGGCCCGCGAGCCGCCGTGCAT 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39231 GTCCGACCGAGCGAGCGCGCTGCTTCGAGGACAGCGCTCCGCGAGACGTCGCGCGAG 39290
    || ||||| ||| | | | | | | | | | | | | | | | | | | | | |
QY 61 CGGCCGCTGTGCGACTGGGCGCTCGACTCGCTACGCGCGT 101
    || ||||| ||| | | | | | | | | | | | | | | | | | | | | |
Db 39291 CGGCGTTCAAGAGCGTGGGCTTCGACTCCCTCACCGCTGT 39331
```

```
Search completed: November 6, 2002, 16:02:52
Job time : 217.625 secs
```


	C	28	35.4	35.0	4403765	22	AAT99683	Mycobacterium tube
	C	29	33.8	33.5	11220	21	AAZ87298	S. venezuelae macr
		30	33.8	33.5	13842	21	AAZ87297	S. venezuelae macr
		31	33.8	33.5	33529	23	AAAS17367	DNA sequence of S.
		32	33.4	33.1	47981	22	AAF30757	Micromonospora meg
		33	32.6	32.3	30690	21	AAA92301	S. avermitilis ave
		34	32.6	32.3	30690	22	AAH79277	Streptomyces averm
		35	32.2	31.9	11916	22	AAH79279	Streptomyces averm
		36	32.2	31.9	12381	21	AAZ58381	Streptomyces averm
		37	32.2	31.9	29879	14	AAQA6806	erya region of S.
		38	32.2	31.9	31422	21	AAA92302	S. avermitilis ave
		39	32.2	31.9	31422	22	AAH79278	Streptomyces averm
		40	31.4	31.1	2811	21	AAAC55794	Mitomycin biosynth
		41	31.4	31.1	5481	22	AAFS1658	Mycobacterium tube
		42	31.4	31.1	5485	22	AAHS2061	Mycobacterium tube
		43	31.4	31.1	18034	21	AAAC5841	Complete Mitomycin
		44	31.2	30.9	1770	24	ABO90065	M. capsulatus gene
	C	45	31.2	30.9	4411529	22	AAI99682	Mycobacterium tube

ALIGNMENTS

```

RESULT_1
ID      AAA29349 standard; DNA; 71989 BP.
XX      AAA29349;
AC
XX      AAAG9349;
DT      12-SEP-2000    (first entry)
XX
DE       Sorangium cellulosum epothillone polyketide synthase operon genomic DNA.
XX
KW       Epothillone; polyketide synthase; epoA; epoB; epoc; epod; epoE; epoF;
KM       epopL; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KN       tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS       Sorangium cellulosum.
XX
FH       Key location/Qualifiers
FT       CDS          3..992
FT              /*tag= a
FT              /label= ORF_A
FT              /product= transposase
FT              /note= "not part of the PKS"
FT              989..1501
FT              CDS
FT              /*tag= b
FT              /label= ORF_B
FT              /product= transposase
FT              /note= "not part of the PKS"
FT              1998..6263
FT              /*tag= c
FT              /label= epoa_gene
FT              /note= "encodes the loading domain"
FT              2031..3548
FT      misc_RNA

```

```

FT      /tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5610
FT      /tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /tag= x
FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15736
FT      /tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /tag= ao
FT      /note= "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /tag= au

```

```

FT      /note= "encodes DH6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes KR6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= epoE_gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      41341..41922
FT      /tag= bd
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No.5.5e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGCAGGCCGAGATCGCGCGCTGCTTCATGAGGGCGCCGACGCCGTGCCGTCGAT 60
Db      31170 GTGCAGGCCGAGATCGCGCGCTGCTTCATGAGGGCGCCGACGCCGTGCCGTCGAT 31229
QY      61 CGGCCGCTGTCGACTGGCGCTCGACTCGCTCACGGCGGT 101
Db      31230 CGGCCGCTGTCGACTGGCGCTCGACTCGCTCACGGCGGT 31270

RESULT 2
AAZ55887
ID      AAZ55887 standard; DNA: 68750 BP.
AC      AAZ55887;
XX      10-APR-2000 (first entry)
DE      Sorangium cellulosum 68.75 kb contig.
XX      Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KM      anticancer; ds.
XX      Sorangium cellulosum.
OS      Location/Qualifiers
FH      1..1826
FT      /tag= a
FT      /partial
FT      /product= "Partial Orf 1 protein (AAV58580)"
FT      /note= "No initiation codon given in the specification"
FT      complement (1900..3171)
FT      /tag= b
FT      /product= "Orf 2 protein (AAV58581)"
FT      3415..5556
FT      /tag= c
FT      /product= "Orf 3 protein (AAV58582)"
FT      complement (5612..5992)
FT      /tag= d
FT      /product= "Orf 4 protein (AAV58583)"
FT      6226..6675
FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
FT      /tag= f

```

```

FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /partial
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
FT      WO9966028-A2.
FT      23-DEC-1999.
FT      16-JUN-1999; 99WO-EP04171.
FT      18-JUN-1998; 98US-0099504.
FT      24-SEP-1998; 98US-0101631.
FT      05-FEB-1999; 99US-0118906.
FT      (NOVS ) NOVARTIS AG.
FT      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GRS MBH.
FT      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;

```

```

XX  WPI: 2000-097741/08.
DR  P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR  AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR  AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR  AAY58592, AAY58593, AAY58594.
XX
XX  New isolated epothilone synthase genes, used for the recombinant
PT  production of epothilone for use in cancer therapy.
XX
PS  Claim 14: Page 87-104; 174pp: English.
XX
XX  This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC  comprising 22 open reading frames (ORFs) and includes genes encoding
CC  proteins involved in the biosynthesis of epothilones. Epothilones A and
CC  B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC  starter unit; polyketides being synthesised from two-carbon building
CC  blocks, the beta-carbon of which always carries a keto group. Each round
CC  of two-carbon addition is carried out by a complex of enzymes known as
CC  the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC  EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC  the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC  and EPOS E (AAY58575-758578) are involved in polyketide backbone
CC  formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC  the protein Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC  involved in transport. Epothilones mimic the biological activity of
CC  taxol, and may be substituted for taxol in cancer chemotherapeutic
CC  compositions. Epothilones exhibit a much lower drop in potency against a
CC  multiply drug-resistant cell line compared with taxol, and are
CC  considerably less efficiently exported from such cells by the multidrug
CC  resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC  epothilones as anticancer agents, they are problematical to produce on a
CC  large scale. Epothilones are too complex for industrial scale chemical
CC  synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC  poor yields of epothilones. The nucleic acids of the invention may be
CC  used for the recombinant production of epothilones in a heterologous host
CC  that is more amenable to fermentation.
XX
SQ  Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other:
XX
XX  Query Match          98.4%; Score 99.4; DB 21; Length 68750;
XX  Best Local Similarity 99.0%; Pred. No. 1.4e-17;
XX  Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX  1 GTGCAGCCGAGATCGCGCGCGCTTTCATGAGCGCGCGCGCGCGCGCTGCAT 60
XX  |||||||
XX  Db 36782 GTGCAGCGCGAGATCGCGCGCGCTTTCATGAGCGCGCGCGCGCGCGCTGCAT 36841
XX
XX  61 CGCGCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101
XX  |||||||
XX  Db 36842 CGCGCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 36882
XX
XX  RESULT 3
XX  AAY21187
XX  ID AAY21187 standard; DNA; 53789 BP.
XX
XX  AAY21187;
XX
XX  24-JUL-1998 (first entry)
XX
XX  Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
DE  Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
XX  Amycolatopsis mediterranei rifamycin synthesis gene cluster;
KM  polyketide synthase; actinomycete; ansamycin; ds.
XX
XX  Amycolatopsis mediterranei.
OS
XX
XX  Key Location/Qualifiers
XX  CDS 1825..15543
XX  /tag= a
XX  /label= ORF_A
XX  /product= "polyketide synthase"
XX

```

```

XX  CDS 15550..30759
XX  /tag= b
XX  /label= ORF_B
XX  /product= "polyketide synthase"
XX  30895..36060
XX  /tag= c
XX  /label= ORF_C
XX  /product= "polyketide synthase"
XX  36259..41325
XX  /tag= d
XX  /label= ORF_D
XX  /product= "polyketide synthase"
XX  41373..51614
XX  /tag= e
XX  /label= ORF_E
XX  /product= "polyketide synthase"
XX  51713..5293
XX  /tag= f
XX  /label= ORF_F
XX  /product= "polyketide synthase"
XX
XX  WO9807868-A1.
XX
XX  26-FEB-1998.
XX
XX  18-AUG-1997; 97WO-EP04495.
XX
XX  20-AUG-1996; 96EP-0810551.
XX
XX  (NOVS ) NOVARTIS AG.
XX
XX  Engel N, Schupp T, Toupet C;
XX
XX  WPI: 1998-169172/15.
XX  P-PSDB: AAY52845-W52850.
XX
XX  Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT  to produce rifamycin and rifamycin analogues
XX
XX
PS  Claim 4: Page 53-102; 205pp: English.
XX
XX  The present sequence represents a Amycolatopsis mediterranei rifamycin
CC  synthesis gene cluster DNA fragment from the present invention. The
CC  DNA fragment comprises a DNA region involved directly or indirectly
CC  in the gene cluster responsible for rifamycin synthesis, including
CC  the adjacent DNA regions to the right and left which, by reason of
CC  their function in connection with rifamycin biosynthesis, qualify
CC  as constituents of this rifamycin gene cluster, and functional
CC  fragments, derivatives or constituents of these. The Amycolatopsis
CC  mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC  for producing rifamycin, rifamycin analogues or precursors. It can also
CC  be used for inactivating or modifying genes involved in ansamycin or
CC  rifamycin biosynthesis. The DNA can be used for constructing mutant
CC  actinomycetes strains from which the natural rifamycin or ansamycin
CC  biosynthesis gene cluster has been partly or completely deleted. The
CC  DNA fragment can be used for assembling a library of polyketide
CC  synthases, which can be used for assembling a library of polyketides.
CC  A hybridisation probe of the invention can be used for identifying DNA
CC  fragments involved in the biosynthesis of ansamycins.
XX
XX
SQ  Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other:
XX
XX  Query Match          47.7%; Score 48.2; DB 19; Length 53789;
XX  Best Local Similarity 67.3%; Pred. No. 0.00042;
XX  Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
XX  1 GTGCAGCCGAGATCGCGCGCGCTTTCATGAGCGCGCGCGCGCGCTGCAT 60
XX  || ||| || ||||| ||||| || || ||||| ||||| || |||||
XX  Db 25279 GTCCGGGCGCAGATCGCGCGCGCTTTCATGAGCGCGCGCGCGCGCTGCAT 25338
XX
XX  61 CGCGCGCTGTGCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101
XX  || ||| || ||||| ||||| || ||||| || |||||
XX  Db 25339 ACGGCGCTTCAAGAGCGCGCGCTTTCATGAGCGCGCGCGGT 25379
XX

```

RESULT 4

AAFB8339

ID AAFB8339 standard; DNA: 16767 BP.

AC AAFB8339;

DT 28-AUG-2001 (first entry)

DE S. spinosa DNA fragment encoding ORF22, SEQ ID 49.

XX Forosamine; trimethylhamose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KM macrolide; insecticidal; polyketide synthase; ds.

XX Saccharopolyspora spinosa.

OS DEL9957268-A1.

XX 08-MAR-2001.

PD 29-NOV-1999; 99DE-1057268.

PF 27-AUG-1999; 99DE-1040596.

PR (FARB) BAYER AG.

PA Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

PI WPI; 2001-267102/28.

DR P-PSDB; AAB70969.

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives

PS Claim 7a; Page 284-313; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (I) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (II) to generate a library of polyketide synthases; (III) for
CC adding forosamine or trimethylhamose to a spinosyn or polyketide
CC aglycone; and (IV) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 22, encodes an S. spinosa polyketide synthase.

XX Sequence 16767 BP; 2256 A; 4805 C; 6457 G; 3249 T; 0 other;

Query Match 40.0%; Score 40.4; DB 22; Length 16767;

Best Local Similarity 63.3%; Pred. No. 0.049; Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGAGGCCGAGATCGCGCGCTTTTCATGGGGCCCGGAGCGCCGTCGAT 60

DB 15661 GTCCGCCCGCATATCGCGGTGACTGGGGCAGCGAGTGTGACATCGAG 15720

QY 61 CGCGCGCTGTGCGACTTGGCGCTGACTCGCTCAGCGC 98

DB 15721 AAGCTTTGGCGAGCTGGGTTTGCACCTCGCTGACGCG 15758

RESULT 5

AAFB8314

ID AAFB8314 standard; DNA: 25360 BP.

XX AAFB8314;

DT 28-AUG-2001 (first entry)

DE S. spinosa DNA fragment SEQ ID 3.

XX Forosamine; trimethylhamose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KM macrolide; insecticidal; ds.

XX Saccharopolyspora spinosa.

OS DEL9957268-A1.

XX 08-MAR-2001.

PF 29-NOV-1999; 99DE-1057268.

PR 27-AUG-1999; 99DE-1040596.

PA (FARB) BAYER AG.

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

DR WPI; 2001-267102/28.

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives

PS Claim 7; Page 49-58; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (I) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (II) to generate a library of polyketide synthases; (III) for
CC adding forosamine or trimethylhamose to a spinosyn or polyketide
CC aglycone; and (IV) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine, trimethylhamose and polyketide synthase biosynthesis.

XX Sequence 25360 BP; 3832 A; 9143 C; 8354 G; 4031 T; 0 other;

Query Match 40.0%; Score 40.4; DB 22; Length 25360;

Best Local Similarity 63.3%; Pred. No. 0.048; Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 GTGAGGCCGAGATCGCGCGCTTTTCATGGGGCCCGGAGCGCCGTCGAT 60

DB 92 GTCCGCCCGCATATCGCGGTGACTGGGGCAGCGAGTGTGACATCGAG 151

QY 61 CGCGCGCTGTGCGACTTGGCGCTGACTCGCTCAGCGC 98

DB 152 AAGCTTTGGCGAGCTGGGTTTGCACCTCGCTGACGCG 189

RESULT 6

AAFB8317

ID AAFB8317 standard; DNA: 29736 BP.

XX AAFB8317;

```

XX 28-AUG-2001 (first entry)
XX
XX S. spinosa DNA fragment SEQ ID 6.
DE
XX Forsamine: trimethylrhinmose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
XX
XX Saccharopolyspora spinosa.
XX
XX DE19957268-A1.
XX
XX 08-MAR-2001.
XX
XX 29-NOV-1999; 99DE-1057268.
XX
XX 27-AUG-1999; 99DE-1040596.
XX
XX (FARB ) BAYER AG.
XX
XX Eberz G, Moehrl V, Froede R, Velten R, Salas JA;
XX
XX WPI: 2001-267102/28.
XX
XX
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives
XX
XX Claim 7; Page 92-102; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactivate or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylrhinmose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine and trimethylrhinmose biosynthesis.
XX
XX Sequence 29736 BP; 4401 A; 10346 G; 10080 G; 4909 T; 0 other;
SQ
Query Match 40.0%; Score 40.4; DB 22; Length 29736;
Best Local Similarity 63.3%; Pred. No. 0.048; Mismatches 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 36; Indels 0;
QY 1 GTGCAGCCGAGATCGCGCGTGTTCATGGGCGCGCGAGCCCGTCCGTCGAT 60
DB 4468 GTCCGCGCGGATATCGCGGCTACTGGGGCAGCGAGTCGATGTGACATCGAG 4527
QY 61 CGGCCGCTGTGGACTTGGGCTCGACTCGCTACGCGC 98
DB 4528 AAGCCTTGTGGCGGCTGGGTTTCGACTCGCTGACGCGC 4565

```

```

DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
XX
XX Spinosyn biosynthetic enzyme: open reading frame; ORF; insecticidal
XX macrolides; arachnid; insect; polyketide; polyketide synthase;
XX PKS; extender module; initiator module; acyl transferase domain; AT;
XX acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
XX dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
XX insecticide; ss.
XX
XX Saccharopolyspora spinosa.
XX
XX
XX Key Location/Qualifiers
XX CDS complement (1135..1971)
XX /tag= a
XX /product= ORFL16
XX /note= "Protein involved in transcription control"
XX 2024..2791
XX /tag= b
XX /product= ORFL15
XX /note= "keto acyl reductase"
XX complement (3416..4165)
XX /tag= c
XX /product= spns
XX /note= "Spinosyn biosynthesis protein S"
XX complement (4168..5325)
XX /tag= d
XX /product= spnr
XX /note= "Spinosyn biosynthesis protein R"
XX complement (5363..6751)
XX /tag= e
XX /product= spnq
XX /note= "Spinosyn biosynthesis protein Q"
XX 7083..8450
XX /tag= f
XX /product= spnp
XX /note= "Spinosyn biosynthesis protein P"
XX 8967..10427
XX /tag= g
XX /product= spno
XX /note= "Spinosyn biosynthesis protein O"
XX 10436..11434
XX /tag= h
XX /product= spnh
XX /note= "Spinosyn biosynthesis protein N"
XX complement (11530..12492)
XX /tag= i
XX /product= spnm
XX /note= "Spinosyn biosynthesis protein M"
XX complement (12696..13547)
XX /tag= j
XX /product= spnl
XX /note= "Spinosyn biosynthesis protein L"
XX complement (13592..14785)
XX /tag= k
XX /product= spnk
XX /note= "Spinosyn biosynthesis protein K"
XX complement (14799..16418)
XX /tag= l
XX /product= spnj
XX /note= "Spinosyn biosynthesis protein J"
XX 16536..17743
XX /tag= m
XX /product= spni
XX /note= "Spinosyn biosynthesis protein I"
XX complement (17749..18501)
XX /tag= n
XX /product= spnh
XX /note= "Spinosyn biosynthesis protein H"
XX complement (18541..19713)
XX /tag= o
XX /product= spng
XX /note= "Spinosyn biosynthesis protein G"
XX 20168..20995
CDS

```

```

FT      /*tag= p
FT      /product= spnF
FT      /note= "Spinosyn biosynthesis protein F"
FT      2111..28688
FT      CDS
FT      /*tag= q
FT      /product= spnA
FT      /note= "Spinosyn biosynthesis protein A"
FT      /transl_except= (Pos:26940..26942, aa:Pro)
FT      2126..22379
FT      misc_feature
FT      /*tag= r
FT      /note= "Beta-ketosynthase domain (KS1): part of the
FT      initiator module"
FT      22692..23669
FT      misc_feature
FT      /*tag= s
FT      /note= "Acyl transferase domain (AT1): part of the
FT      initiator module"
FT      23793..24041
FT      misc_feature
FT      /*tag= t
FT      /note= "Acyl carrier protein domain (ACP1): part of the
FT      initiator module"
FT      24102..25349
FT      misc_feature
FT      /*tag= u
FT      /note= "Beta-ketosynthase domain (KS1): part of
FT      extender module 1"
FT      25683..26684
FT      misc_feature
FT      /*tag= v
FT      /note= "Acyl transferase domain (AT1): part of
FT      extender module 1"
FT      27582..28121
FT      misc_feature
FT      /*tag= w
FT      /note= "Beta-ketoreductase domain (KRL): part of
FT      extender module 1"
FT      28404..28649
FT      misc_feature
FT      /*tag= x
FT      /note= "Acyl carrier protein domain (ACP1): part of
FT      extender module 1"
FT      38916..35374
FT      CDS
FT      /*tag= y
FT      /product= spnC
FT      /note= "Spinosyn biosynthesis protein C"
FT      29024..30295
FT      misc_feature
FT      /*tag= z
FT      /note= "Beta-ketosynthase domain (KS2): part of
FT      extender module 2"
FT      30629..31621
FT      misc_feature
FT      /*tag= aa
FT      /note= "Acyl transferase domain (AT2): part of
FT      extender module 2"
FT      31697..32254
FT      misc_feature
FT      /*tag= ab
FT      /note= "Dehydratase domain (DH2): part of extender
FT      module 2"
FT      33035..34072
FT      misc_feature
FT      /*tag= ac
FT      /note= "Enoyl reductase domain (ER2): part of
FT      extender module 2"
FT      34082..34621
FT      misc_feature
FT      /*tag= ad
FT      /note= "Beta-ketoreductase domain (KR2): part of
FT      extender module 2"
FT      34886..30295
FT      misc_feature
FT      /*tag= ae
FT      /note= "Acyl carrier protein domain (ACP2): part of
FT      extender module 2"
FT      35419..44931
FT      CDS
FT      /*tag= af
FT      /product= spnC
FT      /note= "Spinosyn biosynthesis protein C"
FT      35518..36786
FT      misc_feature
FT      /*tag= ag
FT      /note= "Beta-ketosynthase domain (KS3): part of
FT      extender module 3"
FT      37108..38097
FT      misc_feature

```

```

FT      /*tag= ah
FT      /note= "Acyl transferase domain (AT3): part of
FT      extender module 3"
FT      38992..39528
FT      misc_feature
FT      /*tag= ai
FT      /note= "Beta-ketoreductase domain (KR3): part of
FT      extender module 3"
FT      39790..40035
FT      misc_feature
FT      /*tag= aj
FT      /note= "Acyl carrier protein domain (ACP3): part of
FT      extender module 3"
FT      40102..41373
FT      misc_feature
FT      /*tag= ak
FT      /note= "Beta-ketosynthase domain (KS4): part of
FT      extender module 4"
FT      41713..42705
FT      misc_feature
FT      /*tag= al
FT      /note= "Acyl transferase domain (AT4): part of
FT      extender module 4"
FT      43615..44157
FT      misc_feature
FT      /*tag= am
FT      /note= "Beta-ketoreductase domain (KR4): part of
FT      extender module 4"
FT      44431..44676
FT      misc_feature
FT      /*tag= an
FT      /note= "Acyl carrier protein domain (ACP4): part of
FT      extender module 4"
FT      44966..59752
FT      CDS
FT      /*tag= ao
FT      /product= spnD
FT      /note= "Spinosyn biosynthesis protein D"
FT      45077..46348
FT      misc_feature
FT      /*tag= ap
FT      /note= "Beta-ketosynthase domain (KS5): part of
FT      extender module 5"
FT      46691..47674
FT      misc_feature
FT      /*tag= aq
FT      /note= "Acyl transferase domain (AT5): part of
FT      extender module 5"
FT      47753..48310
FT      misc_feature
FT      /*tag= ar
FT      /note= "Dehydratase domain (DH5): part of extender
FT      module 5"
FT      49226..49771
FT      misc_feature
FT      /*tag= as
FT      /note= "Beta-ketoreductase domain (KR5): part of
FT      extender module 5"
FT      50009..50254
FT      misc_feature

```

```

Query Match      40.0%; Score 40.4; DB 20; Length 80161;
Best Local Similarity 63.3%; Pred. No. 0.047;
Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY      1  GTGCAGCGCGAGATCGCGCGCTTTCATGCGCGCGCGCGCGCGCGCGCGAT 60
DB      75463  GTCCGCGCGCGARATCGCGGTGACTGCGCGCGCGCGCGAGTGCAGATGCGAG 75522

OY      61  CGCCGCTGTGCGACTTGCGCCCTGCATCGCTACGCGC 98
DB      75523  AAGCCTTGCGCGAGCTGGTTTGCATCGCTGACGCGC 75560

RESULT 8
AAA89135/c
ID      AAA89135 standard; DNA; 561 BP.
XX
AC      AAA89135;
XX
DT      19-MAR-2001 (first entry)
XX
DE      Polyketide synthase derived sequence 34-119.1.T7 DNA.
XX
KW      Polyketide synthase; PKS; cosmid 34-119; FR-520; antibiotic;

```


DB 2911 CGCCCGTCGCGAGATCGGCTTCGACTGCGTACCGCGT 2951
RESULT 10
ID AAD17185 standard; DNA; 27541 BP.
AC AAD17185;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KM Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KM antifungal; antibiotic; nys2; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (454..1191)
FT /*tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
FT complement (1275..3092)
FT /*tag= b
FT /product= "NysG protein"
FT complement (3070..4824)
FT /*tag= c
FT /note= "CDS does not include start codon"
FT /product= "NysH protein"
FT 5122..6156
FT /*tag= d
FT /product= "NysD3 protein"
FT 6338..27541
FT /*tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNFE) SINTEF STIPELSEN IMD TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVILK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AF;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR MPI: 2001-557614/62.
DR P-PSDB: AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 2: Page 151-166; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;
XX
Query Match 38.2%; Score 38.6; DB 22; Length 27541;
Best local Similarity 61.4%; Pred. No. 0.14;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 1 GTCCAGCCGAGATCGCGCGCTTTCATGSGGCGCGGAGCGCGCTTCGAT 60
DB 16145 GTCCGACCGAGTCCGCCACCGCTTCGCGCCACAGCTCCCGACCGCGGAGGCG 16204
OY 61 CGGCGCGTTCGCGACTTGGGCGCTCGACTGCGTCAACGCGGT 101
DB 16205 CGCGCTTCGCGAGCTGCGCTTCGACTGCTGACCGCGGT 16245
RESULT 11
ID AA287318 standard; DNA; 36778 BP.
XX
AC AA287318;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM chronic obstructive pulmonary disease; respiratory inflammation;
KM hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 1742..15583
FT /*tag= a
FT /product= "pik gene cluster protein #1 (AAV77200)"
FT 15688..26907
FT /*tag= b
FT /product= "pik gene cluster protein #2 (AAV77201)"
FT 26991..31679
FT /*tag= c
FT /product= "pik gene cluster protein #3 (AAV77202)"
FT 31782..35822
FT /*tag= d
FT /product= "pik gene cluster protein #4 (AAV77203)"
FT 35819..36664
FT /*tag= e
FT /product= "pik gene cluster protein #4 (AAV80997)"
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MING) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR MPI: 2000-160679/14.
DR P-PSDB: AAV77200, AAV77201, AAV77202, AAV77203, AAV80997.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31; 438pp; English.
XX

CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC Streptomyces antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macroclides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthases may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 31.

XX Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;

SO Query Match

Best Local Similarity 38.2%; Score 38.6; DB 21; Length 36778;

Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCGCGCTTCATGAGGCGCCCGGACGCCGCTCCGCTGAT 60

Db 32632 GTGATGCGGCGAGCGGCTCCGTCGCGGACATCGCCGAAGATCCCGTCGAC 32691

Qy 61 CGCGCGCTGTCGACTTGGGCGCTCGACTCGCTACGGCGGT 101

Db 32692 CGCCCGCTGGGAGATCGGCTTCGACTCGCTGACCGCGT 32732

RESULT 12

AA287285 standard; DNA: 37948 BP.

XX AA287285;

DT 05-JUN-2000 (first entry)

DE S. venezuelae plk (macroide biosynthesis) gene cluster, seq ID NO:5.

KM Desosamine biosynthesis; macroide; polyketide; methymycin; pikromycin;

KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;

KM chronic obstructive pulmonary disease; respiratory inflammation;

KM hypercholesterolemia; crop protection agent; ds.

OS Streptomyces venezuelae ATCC15439.

PN W0200000620-A2.

PD 06-JAN-2000.

PF 25-JUN-1999; 99MO-US14398.

PR 26-JUN-1998; 98US-0105537.

PA (MINU) UNIV MINNESOTA.

PI Sherman DH, Liu H, Xue Y, Zhao L;

DR WPI; 2000-160679/14.
DR P-PSDB; AAY77180.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.,
XX synthesis of methymycin and pikromycin -
XX
PS Claim 13; Page 299-315; 438pp; English.

CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC Streptomyces antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macroclides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthases may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.

XX Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

SO Query Match

Best Local Similarity 38.2%; Score 38.6; DB 21; Length 37948;

Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCGCGCTTCATGAGGCGCCCGGACGCCGCTCCGCTGAT 60

Db 35802 GTGATGCGGCGAGCGGCTCCGTCGCGGACATCGCCGAAGATCCCGTCGAC 35861

Qy 61 CGCGCGCTGTCGACTTGGGCGCTCGACTCGCTACGGCGGT 101

Db 35862 CGCCCGCTGGGAGATCGGCTTCGACTCGCTGACCGCGT 35902

RESULT 13

AAA75633 standard; DNA: 38506 BP.

XX AAA75633;

DT 22-JAN-2001 (first entry)

DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.

KM Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;

KM antibiotic; C12-hydroxylase; plck; desosamine biosynthesis;

KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;

KM pikromycin biosynthesis; ss.

OS Streptomyces venezuelae.

PN US6117659-A.

PD 12-SEP-2000.

PF 27-MAY-1999; 99US-0320878.

PR	28-MAY-1998;	98US--0087080.
PR	22-SEP-1998;	98US-0100880.
PR	08-FEB-1999;	99US-0119139.
PR	20-MAY-1999;	99US-0134990.
PR	30-APR-1997;	97US-0846247.
PR	06-MAY-1998;	98US-0073538.
PR	28-AUG-1998;	98US-0141908.
PA	(KOSA-) KOSAN BIOSCIENCES INC.	
PI	Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;	
XX	WPI: 2000-610844/58.	
DR		
PT	New recombinant pick hydroxylase gene of Streptomyces venezuelae useful	
PT	for converting ketolides to antibiotics and as antibiotics and	
PT	intermediates in the synthesis of compounds with pharmaceutical value	
PS		
XX	Disclosure: Columns 15-32; 117pp; English.	
XX		
CC	The present sequence is used to produce the recombinant DNA compounds	
CC	of the invention. The specification describes a recombinant DNA compound	
CC	expressing recombinant polyketide synthase genes in host cells for the	
CC	production of narbonolide, narbonolide derivatives and polyketides that	
CC	are useful as antibiotics and as intermediates in the synthesis of	
CC	compounds with pharmaceutical value. The DNA compounds may also encode	
CC	a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl	
CC	transferase enzymes (useful for conversion of ketolides to antibiotics),	
CC	and the beta-glucosidase enzyme (involved in picrotoxin biosynthesis).	
CC	These compounds are also useful for increasing the antibiotic activity	
CC	of a compound relative to the unhydroxylated compound. The recombinant	
CC	host cells are useful as genetic systems that allow rapid engineering	
CC	of the narbonolide polyketide synthase. These would be valuable for	
CC	creating novel ketolide analogs for pharmaceutical applications.	
XX		
SO	Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other:	
	Query Match 38.2%; Score 38.6; DB 21; Length 38506;	
	Best Local Similarity 61.4%; Pred. No. 0.14;	
	Matches 62; Conservative 0; Mismatches 99; Indels 0; Gaps 0;	
OY	1 GTGACGCGCGAGATCGCGCGCTGCTTCATGAGGGCGCGAGCGCGCCCGCTGAT 60	
Db	32774 GTGATGCGGACAGCGCGCTCCCTGCTCGCGTCCGACTCGACCCGAGAGGTCGCCGCGAC 32833	
OY	61 CGGCGCGTGTGGACTTGGGCGCTGCAGTCGCTCAGCGGCGGT 101	
Db	32834 CGCCCGCTGCGGGAGATCGGCTTCACCTCGCTGACCGCGCT 32874	
RESULT 14		
ID	AAZ56001 standard; DNA: 38506 BP.	
XX	AAZ56001;	
AC		
XX		
DT	23-MAR-2000 (first entry)	
XX		
DE	Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.	
XX		
KW	Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;	
KW	antibiotic production; narbomycin; picrotoxin; ds.	
XX		
OS	Streptomyces venezuelae.	
XX		
Key	Location/Qualifiers	
FT	70..13725	
FT	/tag= a	
FT	/product= PICAT	
FT	/note= "Narbonolide synthase subunit 1"	
FT	13830..25049	
FT	/tag= b	
FT		
CDS		

FT	/product- PICAI1
FT	/note="Narbonolide synthase subunit 2"
FT	25133..29821
FT	/tag- c
FT	/product- PICAI11
FT	/note="Narbonolide synthase subunit 3"
FT	29924..33964
FT	/tag- d
FT	/product- PICAI1V
FT	/note="Narbonolide synthase subunit 4"
FT	33961..34806
FT	/tag- e
FT	/product- PICB
FT	/note="Contains typeII thioesterase domain"
FT	34863..36011
FT	/tag- f
FT	/product- PICC11
FT	/note="4-keto-6-deoxyglucose isomerase"
FT	36159..37439
FT	/tag- g
FT	/product- PICC111
FT	/note="Desosaminyl transferase"
FT	37529..38242
FT	/tag- h
FT	/product- PICCV1
FT	/note="3-amino dimethyltransferase"
PN	WO9961599-A2.
XX	
PD	02-DEC-1999.
XX	
PE	27-MAY-1999; 99WO-0511814.
XX	
PR	28-MAY-1998; 98US-0087080.
PR	28-AUG-1998; 98US-0141908.
PR	22-SEP-1998; 98US-0100980.
PR	08-FEB-1999; 99US-0119139.
XX	
PA	(KOSA-) KOSAN BIOSCIENCES INC.
XX	
PI	Ashley G, Beliach MC, Beliach M, McDaniel R, Tang L;
DR	WPI: 2000-072618/06.
DR	P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
XX	AAY67208, AAY67211.
PT	New recombinant DNA encoding a domain of narbonolide polyketide
PT	synthase, for production of ketolide antibiotics -
XX	
PS	Example 2; Page 16-27; 98pp; English.
XX	
CC	This is the recombinant cosmid pKOS023-27 DNA sequence which contains a
CC	Streptomycetes venezuelae DNA insert. The cosmid contains open reading
CC	frames which encode the various modules of the narbonolide polyketide
CC	synthase (PKS). The invention relates to recombinant DNA containing a
CC	coding sequence for a narbonolide PKS. Polyketides are compounds
CC	synthesised from 2-carbon units through a series of condensations and
CC	subsequent modifications. Modular PKSs are responsible for the production
CC	of many antibiotics including picromycin. The narbonolide PKS consists of
CC	a loading module, six extender modules, and two thioester domains. Four
CC	proteins make up the narbonolide PKS (PICAI, PICAI1, PICAI11 and PICAI1V).
CC	PICAI1 includes the loading module and extender modules 1 and 2, PICAI1
CC	includes extender modules 3 and 4, PICAI11 includes extender module 5 and
CC	PICAI1V includes extender module 6 and a type II thioesterase domain. The
CC	second type II thioesterase domain is found on the PICB protein. The
CC	nucleotide sequences encoding all of these proteins can be isolated in
CC	recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is
CC	desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC	transferase enzyme is required for this conversion, and the desosamine
CC	biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC	DNA of the invention is used to express, in transformed cells,
CC	narbonolide (or its derivatives) or other ketolides (particularly
CC	hybrids), which may then be converted (e.g. by other enzymes

CC recombinantly expressed in the same hosts) to polyketide antibiotics or
CC their intermediates. The antibiotics are useful in human or veterinary
CC medicine.

SO Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 38.2%; Score 38.6; DB 21; Length 38506;

Best Local Similarity 61.4%; Pred. No. 0.14;

Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 GTGCAGCGCGAGATCGCGCGCTTTTCATGGGGCGCGCGCGCGCTGCCCTGCAT 60

DB 32774 GTGATCGCGGACGCGCGCTCCGCTCCGCTGCAGCTCCCGAGAGGTCGCCGTGCAC 32833

OY 61 CGCGCGCTGTCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101

DB 32834 CGCGCGCTCGGAGATCGGCTTCGACTCGCTGACCGCGCT 32874

RESULT 15

AAA09469 ID AAA09469 standard; DNA; 50937 BP.

AC AAA09469;

DT 29-AUG-2000 (first entry)

DE Streptococcus oleandomycin gene cluster.

KW oleandomycin; oleandrolide; polyketide synthase; oleA1; oleA11; oleA111;

KW PKS; type I; 8,8a-deoxyoleandrolide synthase; modular; ketosynthase;

KW acyl-transferase; acyl carrier protein; inactivated; polyketide;

KW macrolactone; antibiotic; mollide; erythromycin; ss.

OS Streptococcus antibioticus.

Key Location/Qualifiers

FT 152..1426

FT /tag= a

FT /label= ole1

FT complement (1528..2637)

FT /tag= b

FT /label= olen2

FT complement (2658..4967)

FT /tag= c

FT /label= oler

FT 5772..18224

FT /tag= d

FT /label= ORF1

FT product= 8,8a-deoxyoleandrolide_synthase_1

FT 18267..29717

FT /tag= e

FT /label= ORF2

FT product= 8,8a-deoxyoleandrolide_synthase_2

FT 29787..40346

FT /tag= f

FT /label= ORF3

FT product= 8,8a-deoxyoleandrolide_synthase_3

FT 40625..41830

FT /tag= g

FT /label= oleP1

FT 41878..43158

FT /tag= h

FT /label= oleG1

FT 43163..44443

FT /tag= i

FT /label= oleG2

FT 44433..45173

FT /tag= j

FT /label= oleM1

FT 45251..46411

FT /tag= k

FT /label= oleX

FT CDS

FT 46491..47714

FT /tag= l

FT /label= oleP

FT complement (47808..49517)

FT /tag= m

FT /label= oleB

FT WO200026349-A2.

FT 11-MAY-2000.

FT 22-OCT-1999; 99WO-US24478.

FT 29-OCT-1998; 98US-0106100.

FT 16-FEB-1999; 99US-0120254.

FT (KOSA-) KOSAN BIOSCIENCES INC.

FT Betlach MC, Shah SK, McDaniel R, Tang L;

FT WPI: 2000-365602/31.

FT P-PSDB: AAY92707, AAY92708, AAY92709.

FT Recombinant DNA compound encoding oleandrolide polyketide synthase for

FT synthesizing polyketides comprising a coding sequence for a domain of a

FT loading module or any one of extender modules

FT Disclousure; Page 14-26; 86pp; English.

FT This is part of the Streptococcus antibioticus oleandomycin gene cluster.

FT The oleandrolide polyketide synthase (PKS), also known as

FT 8,8a-deoxyoleandrolide synthase, is encoded by three open reading frames

FT (ORF), designated oleA1, oleA11 and oleA111. The PKS is a type I

FT "modular" enzyme, where each ORF encodes 2 extender modules and

FT the first ORF also encodes the loading module. Each module is composed

FT of at least a ketosynthase (KS), acyl-transferase (AT) and an

FT acyl carrier protein (ACP) domain. The oleandrolide PKS loading module

FT contains an inactivated KS, called KS-Q, where Q is the abbreviation for

FT glutamine, present instead of the active site cysteine required for

FT activity. The large multifunctional PKS enzymes catalyze the biosynthesis

FT of polyketide macrolactones through multistep pathways involving

FT decarboxylative condensations between acylthioesters followed by cycles

FT of varying beta-carbon processing activities. The macrolide product of

FT the PKS, 8,8a-deoxyoleandrolide, is further modified by epoxidation and

FT glycosylation to yield oleandomycin, an antibacterial polyketide. The

FT invention concerns an isolated recombinant DNA compound, comprising a

FT coding sequence for a domain of loading module or any one of extender

FT modules 1-4 or 1-6, including an oleandrolide PKS operably linked to a

FT promoter. Also discussed are recombinant oleandrolide PKS in which the

FT module 1 KS domain is inactivated by deletion or other mutation. In

FT particular, the inactivation is mediated by a change in the KS domain

FT that renders it incapable of binding substrate (the KS1-o mutation),

FT rendered by mutation in the codon for the active site cysteine. The

FT oleandrolide PKS is useful for synthesizing polyketides, which are useful

FT as antibiotics and mollides. Heterologous expression of oleandrolide PKS

FT in host cells such as Streptomyces coelicolor and S. lividans is also

FT made possible. Unmodified oleandrolide compounds can be provided to

FT CC cultures of Saccharopolyspora erythraea and converted to the

FT corresponding derivatives of erythromycins A-D.

FT Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

FT Query Match 38.2%; Score 38.6; DB 21; Length 50937;

FT Best Local Similarity 61.4%; Pred. No. 0.14;

FT Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 GTGCAGCGCGAGATCGCGCGCTTTTCATGGGGCGCGCGCGCGCTGCCCTGCAT 60

DB 39231 GTCCGTACCGAGAGCGCGCTGTCTGGGGCGCGCGCGCGAGAGTCCCGCGCAG 39290

OY 61 CGCGCGCTGTCGACTTGGGCGCTCGACTCGCTCAGCGCGGT 101

DB 39291 CGCGCGTTCAGGAGCTGGGCTTCGACTCCCTACCGCGGT 39331

Fri Nov 8 09:18:40 2002

us-09-724-876-2_copy_31170_31270.rng

Page 13

Search completed: November 6, 2002, 12:16:36
Job time : 271 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 : Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_31170_31270

Perfect score: 101

Sequence: 1 gtgcagcgccgagatcgcgcg.....ctgcactgcctcagcgagcgt 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: GenBank: *
2: gb_da: *
3: gb_htg: *
4: gb_in: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sy: *
12: gb_sts: *
13: gb_un: *
14: gb_vi: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_com: *
21: em_or: *
22: em_ov: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_un: *
28: em_vi: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_man: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	71989	6	AF172664	AF172664 Sequence
3	99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6	AR193029	AR193029 Sequence
5	99.4	98.4	68750	6	AR199551	AR199551 Sequence
6	99.4	98.4	68750	6	AR199559	AR199559 Sequence
7	99.4	98.4	68750	6	AR199567	AR199567 Sequence
8	99.4	98.4	68750	6	AR201097	AR201097 Sequence
9	99.4	98.4	68750	6	AR208671	AR208671 Sequence
10	48.2	47.7	53784	1	AMM223012	AMM223012 Amycolato
11	48.2	47.7	53789	6	A69720	A69720 Sequence 3
12	48.2	47.7	90445	1	AF040570	AF040570 Amycolato
13	41.8	41.4	82746	1	AF453501	AF453501 Actinosyn
14	40.4	40.0	16767	6	AX089464	AX089464 Sequence
15	40.4	40.0	25360	6	AX089418	AX089418 Sequence
16	40.4	40.0	29736	6	AX089421	AX089421 Sequence
17	40.4	40.0	80161	1	AY007564	AY007564 Saccharop
18	40.4	40.0	80161	6	AR165018	AR165018 Sequence
19	40.2	39.8	84985	1	SM278573	SM278573 Streptomy
20	38.6	38.2	11480	1	STMPK30RF	STMPK30RF Streptomy
21	38.6	38.2	24225	1	SC2C4	SC2C4 Streptomy
22	38.6	38.2	27541	6	AX211706	AX211706 Streptomy
23	38.6	38.2	37948	1	AF079138	AF079138 Streptomy
24	38.6	38.2	50937	6	AR159871	AR159871 Sequence
25	38.6	38.2	78210	1	AB070949	AB070949 Streptomy
26	38.6	38.2	113193	1	AF357202	AF357202 Streptomy
27	38.6	38.2	123580	1	AF263912	AF263912 Streptomy
28	38.6	38.2	125401	6	AX211739	AX211739 Sequence
29	37	36.6	9513	6	AX089460	AX089460 Sequence
30	37	36.6	14775	6	AX089462	AX089462 Sequence
31	37	36.6	20394	1	SM132222	SM132222 Streptomy
32	37	36.6	20394	6	AX067996	AX067996 Sequence
33	37	36.6	43280	1	SFU78289	SFU78289 Streptomy
34	37	36.6	50000	6	AX089417	AX089417 Sequence
35	37	36.6	50000	6	AX089420	AX089420 Sequence
36	37	36.6	104326	1	AB070940	AB070940 Streptomy
37	35.8	35.4	65140	6	AX211705	AX211705 Sequence
38	35.4	35.0	1434	6	AR173224	AR173224 Sequence
39	35.4	35.0	5676	6	A69718	A69718 Sequence 1
40	35.4	35.0	15970	1	AE007061	AE007061 Mycobacte
41	35.4	35.0	27522	1	AB070942	AB070942 Streptomy
42	35.4	35.0	30000	6	AX250261	AX250261 Sequence
43	35.4	35.0	30000	6	AX250263	AX250263 Sequence
44	35.4	35.0	41097	1	AF016585	AF016585 Streptomy
45	35.4	35.0	53450	1	MTV018	MTV018 Mycobacte

ALIGNMENTS

RESULT 1
LOCUS AF217189
DEFINITION AF217189 58733 bp DNA linear BCF 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS Polyanium cellulosum.
SOURCE Polyanium cellulosum
ORGANISM Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyanigaceae; Polyanium.
REFERENCE 1 (bases 1 to 58733)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (3453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source location/Qualifiers
1. 58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1. 992
/note="orfa"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VKGHEAFADFGVPRVLLYDNLKNAVVERGDAITREPTLLAT
SADYREPRVAVARGNEKGRVERAIRYREGFPFARAVADGLNRQATEWTSNAL
DSRWEDRARVRAQFEDDSVLRHPDTPPDHREVEVEGKTPARLDLTSYVPH
DSTRRLVVLADLSQVRIDGNOIVATHYRSMRDGOQIEPHELOLVDEKRRARHR
GDRLARARASQAFIRIYAEERGDVNGSIALRLDLDVGAAELPEALFEVLENDTI
HGAVRQVYDRRKRSEHRLPPVSIPTREHMAVYTPSLTTTDLAKKDP"
989. 1501
/note="orfb"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTETKDRLKSLGFLGLACMEQLADPMLREVALIERE
RHKRSLERLKNSRYAAEFKPMTPDSDMPKIDREAVDOLYDSRYADLLFEVTRRYD
AOKPLLSNKAFAWMGVFPHACVVTLDVRLVRAVIEIEASRYLKEKELNAT
RTKORRTKKH"
1998. 6263
/gene="epoa"
1998. 6263
/gene="epoa"
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRIERAAEDPIAINGACRLPGCVIDISGFWTLLGSDRTV
GOVPAERMDAAWFPDDBADPKCTPTYTRASFSLSDVACFDASFFGTSAPRALMDPAHR
LLEVCWEALENAAIAPSAVLGTETGVFGLIGPSEEAALPRATASAEIDAIGGLGM
PSVGAGRISYVGLRGPCVAVDTAVSSSLAVHALCOSLRSEGSTALAGGLSLMISP
STLWMLSKTRALATDGRCAFAEADGFGRGCAVYVVKRLSGARADDRLLAVIRG
SAIINHDASSGLTVPNGSIOETVLKRALADAGCAASVYGAHGTGLDPTIEOA
LNAVYGLGADVATPLLIGSVKTNLGRVYASGITGLKLYLSLQHQIPAHLHAALN
PRISGDLRLVTRAKTPWDMNTPRRKGVSSFGSGTNAHVLEAPATCTPPAPE
RPAELLVLSARTAAALDAHAARLDHLETPSQCLGDAFSLATTSRAEHLAAVAT
SSEGLRALDAAGCOTPGVVGTIADSSRGRLAFLTGOGAOTLGMGRGLDVWPAF
RAFDLCVRLFNQELDRPLREYMAEPASVDALALDOTAFOTOPALFTFEYALALMRS
WGPELIVAGHSIGELVACVAGVSELEAVFLVARGTLMALPAGGMAVSIAPBEA
DYAAVAAPHAASVSTAANVPDQVIYIAGAGCPVHAIAAAMAARGATKRLHSHAFHS
PLMAPPLEHFGRAVSVSTRPSIVLVSLSGAKGTDEVSFGYVWRHAREVVRADG
VVALHAAGAGTVEVGPKSTLLGLVPACLPDRPALLASRAGRDEPVALIGLGM

AVGLVSNAGLPFSGGRVPLEPTYPWQNERWITDKADDAARGDRAPAGHDEYK
GAVRGDRASRLDHPPEPSGRREVVEAGDPFLELIDEGVDRLVLRVTERAPG
LGEVEIADDAAGLSFNDVOLGAVPDDLPGBNPDVLIDEGECAGRIYAGVNGVLV
VGOPIVATAGGAFATHTTSAALVLRQALASATAPAAAPYALTAALDGLALOP
GERVLIIAATAGVLAAYOMQHVGAETHAATGPEKARVLESIGVAVSDSRSEFV
ADVRAWTGEGVDVYVNLSSGLIDKSNLRSRHPFELCKRCQYANONGIAPRELR
NLSESLVDGRMMERPARVRLFEELGLLIAAGFTPTPLATIDIAVAADFTSMQ
AOHLKLVLTLDGPEVOAIRPHAGAGSTSDRDLDRLDLASAAPARAALAEALRQ
VSOVLRTEPIKGAELFLRIGMDSLAMVEILNRNLEASIKLSTPLSTSPNIALLT
ONLIDLATALSLEHVAENLRAVQSDVFSVAGADQDWETIAL"
6260. 10492
/gene="epob"
6260. 10492
/gene="epob"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINOLNLEHOGVKLAADGERLOIQAAPKMLPNLARISEH
KSTILTMQRPARSIYAPARERHVPPLDIOGSPVLTGRTGATPSPGTHAYREX
CTLDLVARLSRAFRKVVARRHDLRAHTLPDMAQVIERVVDIDETIDRGIDSTREA
RLVSLRDAMSHRIYDTERPPLYHVAVALDQQRVLVLSIDLINVDGLSLIIFKDWL
SEYEDPETSLPYLESLRDYVLALESRKSEAHORGMYMKRVAELPEPMLPKAD
PSTLERIEPRHTEQMLPDSWSRLKORSGEALPTGVYILAAFSVIGRMSAPFEL
NITLENRLVPHRVNDITGDFTSVWLIDIDTRDKSPFQRAKRIQEOLEWAMDHVS
GLENOREARVLGIORGALFPVVLISALNOOVGYSTORIGCTHYVSTQRPOLLIDH
QIEMGDVILAMDIVDVFPPDLDDMLNATVATRLRTEPSSQRCSLPRAQLE
AASANENLSLSEHTLGLFAAREQVLPMLAVASAKTLTELSRSRSLGARLR
BOGANPRLVAVAMERKGOVAVLALESAAVYLPIDADLPAEHLHLDHGEVGLV
LTQWDLKLSMPGPIORFLVSDAIVEDGDSLPMPPIQTSVDLAIVYITGSGTGLPK
GVMDIQRGVNVTIDINERFLEPGDRLDLSLSELDVYDGLVILAACTGIVPDA
SKLRDPAAMALIEBERGVYVNNVSNALMBRHYEHEGSDGLASRLSLSGMIPY
GLPGLQAIRGVYSISGATKASINSITGPVRVNDPSKASITPGRPLRQTRIVLD
EALERPVAWPGQLYITGVGALGLYRQDEETKRSFLVHPETGKRLITGDLGRPLD
GNIEFGRREDNQIKLGRYVELGELEETLKSHPNVADIVPVGDAANKLLLVVYP
EGTRRRAAEODASLKTERRVDARAHAAKADGLSDGRVOFKLARIHGLDGLKPVYLD
TGLVPREAGLDVYARRSRVTFLEAPIFVEFGRPLSCLSSVEPGLALPKFRYPGAG
SNYPQOTYAAVAKSGRIEGVDGEFYVHPFERHLIKVSHGIERGAHVQNDVDEAA
FGLLVGRIDATESIXSLSRECLLEAGYAAQLMEBAPSCNICYCVGQDPDEQVR
PVLDRHSDVYVHGMIGRVPDRQFQVCTLDQDSSPRATTRGAPGMDHAPDLIRQ
FLVRLKLPYMEPTVEVLEDALEPVSNGVKRALERKDTSSPRHSGADPDALEELI
LVAAVREVGLLEVVOLOQSFVDLGATSIHIVMRSLQKRLREIATITELFOYPLGS
LASGVRDSKSLDEOPRNMQDRVEARKGRRS"
10639. 16137
/gene="epoc"
10639. 16137
/gene="epoc"
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEBOSSAIAVIGSGRRPGARDDDEFMRNLROGTEAVORSEBO
ELASAGPAVLDPNVYAGVSGVVEDVDFPAAPFGISPPRAERLDDPHRIFEMCAME
ALENAGCPYPTAEGSIGVYACANSSYITSLNHEPRAMPKPMFOTITGDKRYLAT
HVSYRLNLRGPIISVQACSTSLAVNACSLDFPQEGSQWQMGKGLVIRHGRGVYV
AEGGIFSPDGRARDAKANGITMONGGVVLPFLDALSDGDPVRAVILIGSATND
GARKIGFAPSPVGOAOLAMELALGAENASIOYIETHGTGLGDPIETALAEARVY
DDASTRSCAGISGVYKLESAAGIAGLIKTVLALHEHQLPSPSLNEESPNSIDF
ASSPYYVTSIKDMMWSTGTPRRAGSSFGCTGTAAHVLEBAPAAKLPAAAPASAEI
FVYSKSAALDAAARARLDHQAHQGLSGDVAFSLATTSPIHEHLMAAPASREAL
REGDLAARGQTPPGAVARGSPGVNPPVVEFPQGSQWQMGKGLVIRHGRGVVYV
SACDIAIAEGWMSLAEIADBSQSVLEEDVVOYFALVAFALMRMGVGVV
VIGHMGVGAHAHVAGALSLDEAVAIICRRSLRLRISGCEMAVTELSIAEADAALR
GYEDRSVAANSNSPRSTVLSGEPALICVLSLNAKGVCFGRVAVDVAHSDPOVLR
EDLLAALGLRPRRAAVPMRSTVTCAMYAGCELCANTYMMNLQDPVRAEYVQALOG
GHGLEVNSPHILTTVEEMRAQORGAALVGSIRKQDRPALAEALGLAMQGT
VPMGRLPFAGGRVPLPTYPWQERYWIEAPKSAAGDRAGGHPDLGEMOTLST


```

0TSTRMETTLDEIKRLFWLGDHRVQAVVPGAAVLEMAISGAELGDPLOITTVV
LALALFAGDAAVLVQVYVTTTEQSGRLOPFIASRAPAGASRVHAR3ALRVERE
VAPGLTISAVRARLOASMPAAATYAELTEGLOIPFOGIAELRMBRE3ALRVERE
DAAGSAEYRLHALIDACFOVYGSLEFAGGCAFPWVPGVSRILQORSSELMCHA
RVYNGHQTPDRGADPWWYDSSGAVVAEYGLVQALPCGVRRREDDWFLTELEMP
AAVGTAGLVNAGRWMLLGGGGGLGALRSMLEAGHAAVHAASENTSAAGVALLAKAF
DQGAPTAVVHLGSLDGGELDPGLAGALDAPRASVADALDPALVRCDSVLTVD
QALAGMCFRDPRLMILTLCGAQVAGADVSTQALPGLGRVIAEMEHADRCARVDI
PTPDDGELTALIELLADDAEVALGGRGVARYRQRPETRRRRTSCPTDVT
TRADSTLYLNGLGGGLGSVAGMLAEAGACHLVLVGSSAASVQRAVVALEHGR
VYAKADVADRQDLERLREVTTSGMPLKRVHAGILLDGLMQOTPAFRKVMAR
VQGLALHALTRAPLPSFVLVAVGVLSPGQGNAAANTELDLAHHRKQGLPA
LSYDWGLFAEVGMAAAQEDRGARLVSGMRSLPDEGLSALRELEGARQVQMPVN
PRLMVELYPAASSRMLSLRTAHRAAGAGGAPGDGLRLAAAEPSARSGLEPLLR
AOISOVLRLPEKGLIENDAPLISGMNSIMGLERLNRTEAMGLITVPATILMTYPTVA
LSGHLAREACEAPVESPHTTADSAYIEEMSDDLTQLIAAKFKALT"
16134..37907
/gene="epod"
16134..37907
/gene="epod"
16134..37907
/gene="epod"
/note="type I polyketide synthase: epothilone PKs modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF26904.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQONPLKQAIIITQREERLAGLAQALERTPEIATV
GICRPPGADAPAEWELLDAERDAVQPLDRMALGVAVEVPMALITEPIDC
FDAPFGISPREARSIDPQHRLLLEVMEGLDAGIPRISIDSRTEVFVGAFTADP
RTVAPLRERDAYSATGMSTIAGSLSTLGGGCLVTDAGSSSLVAILHACS
LRAGESSLALAGVSTLSPDMMAARQALSPDGRCFDASAGVSEGGEGIVY
LKRSLDAQRDRIHALIRGSATINHGRSGILRAPVLAETVIREAKRANHAEAV
DYETHGTGTSIDPLEVEALRATVGPARSDDGRCVGAATNIGHLEAAGVAGLIK
AALSLTHERIPRNLNERTLNPRIRLEGSALALATEVPWPRTDRPRAGVSSEGMGT
NAHVYLEAPAVEIMPAPERSAELVLSCGSGALDQAARLEHIDMPELGLGY
AFSLATRSAMTHRLAVALTSREGILALSAVAGOGTPGAARCIASSSGKLLALT
GGGQNPGMGRGLCAAMPARFAEADRCVTLFDEPLDPLEVMAAEVSRILLDOT
AFTOPALFAVEYALITAMRSMGVEPELVHSHSGEYAAACVAGVFSLEDEVRLLAAG
RLMUGLSAGGAMVSLGAEVAAVAAPHAMVSTIAVNVPEQVIVAGVQAVOALIA
GFAARGVTRRLVSHAFSHPLMEPLMEERGVAASVYRRPSVLSVLSNGKVVIDE

Query Match      100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 3,4e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGCAGCGCCGAGATCGCGGCTTTCATAGGGGCGCGGAGCGCGTCCGTCGAT 60
|||||
Db 31170 GTGCAGCGCCGAGATCGCGGCTTTCATAGGGGCGCGGAGCGCGTCCGTCGAT 31229
|||||

OY 61 CGGCCGCTGTGCGACTTGCGCTCAGCTCAGCGCGGT 101
|||||
Db 31230 CGGCCGCTGTGCGACTTGCGCTCAGCTCAGCGCGGT 31270
|||||

RESULT 2
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-Oct-2001;
FEATURES
Location/Qualifiers
1..71989
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

0TSTRMETTLDEIKRLFWLGDHRVQAVVPGAAVLEMAISGAELGDPLOITTVV
LALALFAGDAAVLVQVYVTTTEQSGRLOPFIASRAPAGASRVHAR3ALRVERE
VAPGLTISAVRARLOASMPAAATYAELTEGLOIPFOGIAELRMBRE3ALRVERE
DAAGSAEYRLHALIDACFOVYGSLEFAGGCAFPWVPGVSRILQORSSELMCHA
RVYNGHQTPDRGADPWWYDSSGAVVAEYGLVQALPCGVRRREDDWFLTELEMP
AAVGTAGLVNAGRWMLLGGGGGLGALRSMLEAGHAAVHAASENTSAAGVALLAKAF
DQGAPTAVVHLGSLDGGELDPGLAGALDAPRASVADALDPALVRCDSVLTVD
QALAGMCFRDPRLMILTLCGAQVAGADVSTQALPGLGRVIAEMEHADRCARVDI
PTPDDGELTALIELLADDAEVALGGRGVARYRQRPETRRRRTSCPTDVT
TRADSTLYLNGLGGGLGSVAGMLAEAGACHLVLVGSSAASVQRAVVALEHGR
VYAKADVADRQDLERLREVTTSGMPLKRVHAGILLDGLMQOTPAFRKVMAR
VQGLALHALTRAPLPSFVLVAVGVLSPGQGNAAANTELDLAHHRKQGLPA
LSYDWGLFAEVGMAAAQEDRGARLVSGMRSLPDEGLSALRELEGARQVQMPVN
PRLMVELYPAASSRMLSLRTAHRAAGAGGAPGDGLRLAAAEPSARSGLEPLLR
AOISOVLRLPEKGLIENDAPLISGMNSIMGLERLNRTEAMGLITVPATILMTYPTVA
LSGHLAREACEAPVESPHTTADSAYIEEMSDDLTQLIAAKFKALT"
16134..37907
/gene="epod"
16134..37907
/gene="epod"
16134..37907
/gene="epod"
/note="type I polyketide synthase: epothilone PKs modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF26904.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQONPLKQAIIITQREERLAGLAQALERTPEIATV
GICRPPGADAPAEWELLDAERDAVQPLDRMALGVAVEVPMALITEPIDC
FDAPFGISPREARSIDPQHRLLLEVMEGLDAGIPRISIDSRTEVFVGAFTADP
RTVAPLRERDAYSATGMSTIAGSLSTLGGGCLVTDAGSSSLVAILHACS
LRAGESSLALAGVSTLSPDMMAARQALSPDGRCFDASAGVSEGGEGIVY
LKRSLDAQRDRIHALIRGSATINHGRSGILRAPVLAETVIREAKRANHAEAV
DYETHGTGTSIDPLEVEALRATVGPARSDDGRCVGAATNIGHLEAAGVAGLIK
AALSLTHERIPRNLNERTLNPRIRLEGSALALATEVPWPRTDRPRAGVSSEGMGT
NAHVYLEAPAVEIMPAPERSAELVLSCGSGALDQAARLEHIDMPELGLGY
AFSLATRSAMTHRLAVALTSREGILALSAVAGOGTPGAARCIASSSGKLLALT
GGGQNPGMGRGLCAAMPARFAEADRCVTLFDEPLDPLEVMAAEVSRILLDOT
AFTOPALFAVEYALITAMRSMGVEPELVHSHSGEYAAACVAGVFSLEDEVRLLAAG
RLMUGLSAGGAMVSLGAEVAAVAAPHAMVSTIAVNVPEQVIVAGVQAVOALIA
GFAARGVTRRLVSHAFSHPLMEPLMEERGVAASVYRRPSVLSVLSNGKVVIDE

Query Match      100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 3,4e-12;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGCAGCGCCGAGATCGCGGCTTTCATAGGGGCGCGGAGCGCGTCCGTCGAT 60
|||||
Db 31170 GTGCAGCGCCGAGATCGCGGCTTTCATAGGGGCGCGGAGCGCGTCCGTCGAT 31229
|||||

OY 61 CGGCCGCTGTGCGACTTGCGCTCAGCTCAGCGCGGT 101
|||||
Db 31230 CGGCCGCTGTGCGACTTGCGCTCAGCTCAGCGCGGT 31270
|||||

RESULT 3
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain so ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE .
ORGANISM Polyangium cellulosum.
Polyangium cellulosum.
Bacteria: Proteobacteria: Polyangiaceae: Polyangium.
Myxococcales: Sorangineae: delta subdivision: Myxobacteria:
1 (bases 1 to 68750)
Molnar,I., Schnupp,T., Ono,M., Zirkle,R., Milamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Strathmann,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligdon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)
10662695
MEDLINE 20130945
PUBMED 2 (bases 1 to 68750)
Molnar,I.
JOURNAL Direct Submission
TITLE Submitted (03-DEC-1999) Natural Product Genetics, Novartis
REFERENCE Agricultural Research Institute, Inc., 3054 Cornwalls Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA
JOURNAL Location/Qualifiers
1..68750
FEATURES
SOURCE
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARNTSARVLDGDGRATERBVLAEARIGDLRALREH
LRIOEGSPRHCMDIGLTVELLADHDLPLASIFPHARRSLRHDMWSDALVGLPV
RWLAARAGSPLEIREBERRARTAGQARKYMLAAAPRCAPLPREPEDANLPLGP
MSPEVAERRLRASTVATPELACALALAWLGTGAPSGSPATIMELPELTVIRANGAND
AIAAASAPGTSEALRGARLFAWEVSSKSQLGIPALMERLRTVIRANGAND
LSRPEREALIAAEVRLRAOPAPFAGAGLAVAGSSGRLSGLVTDGALVSGDGD
IMFOPGRISPVVLACTDPEFELAPLSOMLVFAHNAAGTISKVLEGSPLITVAMN
QARPMVLAHAGFMAMVNOAMVDPPREGAPFVORSTIMEFPTPCRCHEPAGSAFS
LACDEHLVYCELSAGLUELMRPHRRGAPSPRAYIGERPIATVTPSTLNATVYL
WADPDRAILGVDRKTVETVLAETHRPHAVVSEDRDIFALTGDPDSKDMVHEIIR
SGASTVADYQRLMDPDMVNLNRGLFTTNRIITLARS"
complement(1900..3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and streptomycetes coelcolor and
to Dd-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
```


Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 36882

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 36882

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 36882

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 36882

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 60
|||||
Db 36782 GTGCAGGCCGAGATCGCCGCGTCTTCATGAGGCGCCGAGCGCCGCTGCAT 36841

Qy 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGGGT 36882

RESULT 8
AR201097 68750 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6358719.
DEFINITION
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 GTGCAGGCGGAGATCGCGCGCTTTCATGGGGCCCGCGAGCGCGCTGCCGTCGAT 60
|||||
Db 36782 GTGCAGGCGGAGATCGCGCGCTTTCATGGAGCCCGCGAGCGCGCTGCCGTCGAT 36641
|||||

Y 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
|||||

RESULT 9
AR208671 68750 bp DNA linear PAT 20-JUN-2002
LOCUS Sequence 1 from patent US 6383787.
DEFINITION
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.5e-12;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 GTGCAGGCGGAGATCGCGCGCTTTCATGGGGCCCGCGAGCGCGCTGCCGTCGAT 60
|||||
Db 36782 GTGCAGGCGGAGATCGCGCGCTTTCATGGAGCCCGCGAGCGCGCTGCCGTCGAT 36641
|||||

Y 61 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 101
|||||
Db 36842 CGGCCGCTGTGCGACTTGGGCTCGACTCGCTCAGCGCGGT 36882
|||||

RESULT 10
AMM223012 53784 bp DNA linear BCT 09-FEB-1998
LOCUS

DEFINITION Amycolatopsis mediterranei genes encoding rifamycin polyketide synthases, ORFs 1 to 5.
ACCESSION AJ223012
VERSION AJ223012.1 GI:2764760
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.
SOURCE Amycolatopsis mediterranei.
ORGANISM Amycolatopsis mediterranei
Bacteria; Actinobacteridae; Actinomycetales; Pseudonocardiaceae; Amycolatopsis.
REFERENCE 1 (bases 1 to 53784)
AUTHORS Schupp,T.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel, SWITZERLAND
REFERENCE 2 (bases 1 to 53784)
AUTHORS Schupp,T., Toupet,C., Engel,N. and Goff,S.
TITLE Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from Amycolatopsis mediterranei
JOURNAL Unpublished
FEATURES
source Location/Qualifiers
1..53784
/organism="Amycolatopsis mediterranei"
/strain="LBG A3136"
/db_xref="taxon:33910"
1336..15543
/function="CoA ligase, acylcarrier protein, b-ketocacyl-ACP-synthase, acyltransferase, ketoreductase"
/note="ORF1: PKS modules 1-3"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CA11035.1"
/db_xref="GI:2764761"
/db_xref="SPTREMBL:O54666"
/translation="MRTDLIKPLVLALENATRRPKPAFADHRTVYTGDEATRR LAGHLGLGVRRHGRDVAICLGNRVSVESYFAILRAGAAGVPLNGSATLEHPLTD SGATVVVTDAAQAARLRLAPHELLVTGDDVEGASHYDELALSPAEPAADLDE PAMFETSGTGRPKGVSTORNLMSVSCVYPPGSLDORVLPMLPLFSLSHIA CYLSATVVGASVRADGSSADVMRLIESSTPLAGVPTTYHHVLRARARVSCFAPS LRTGLAGAVIAGAGRSREDETPVPLIDAYSTSCGATINNNPDAARVESCCLAV PCVDVAVDPDTGLDVPAGEBEVWVSGVNLVGLHNSPEATTAAMRDWFTGLAR RDDAGFYITCGRIKELIRGANGIHGEVEALRVYDGAADAAVAGVPGDITLGEVPA VYIPGPTGDPALILEKREQLSAVKVPDRIEVAHPIPTASGRITRGILDEPAQLR VATEHEBOSRADESVAAALRLARLSGLDERACELLEDLVTRQADVYIGQVPGRA FRDIGFTSLAIVELRRLRLEHNGIMLPASVADHPPLAARVAAELIGLQVAAPV VVAADGEPRIATVGAACRLPGVYASPEDIMRVARVYAVASEFPEDRCMDLSLTD PRRAGSTVGGGFLHDAGEFDGFEGLSPRAVADDPQQRLLLETSMALGNACVDP IALKGTDTGVFSGLMGGGSGAVAPBELEGFTVYTVASSVAGRSVYGLGEPATV DYACSSSLVAMHLAAQALROGECSMALAGGVVMTPGSFVFSRQRLADPGRKAR AAADGTGWESEGVVLERLSVAREGRHRLIAYRGSVANDGASNGLTPANGLSQQ RYIIRLALAAAGLAPSDVYVEAHGTTGDDPLEAOALATYGOBRKOPIMAGSKSN IGHQAAGAGVAGVITKYNQALRHEHTLPRTILHPTLEVDMSGALIELTEARAPRNC RPRKGVSSFGVSGTNAHLITLDEAPREEVYAPLPPVPLVYASSTESLSQARLA SLLEGDSVLTEVAGALVSRRVADERAAVVAASALREAVTGLALMTASGTPGVVW FPGQGTOMAGMGRELALBSPFAERIAECALAPALWIMSLTDVLRGDLGRVDYIQ PACFAGMGALAAWESVGVPRDAVVGVSQGEITAAACVSGALSLEDAAKVALRSGAIA AELSGGASVVALEDGDDVYLVGVGEVPAAVNGSSVYIADDAHALDLETILSSEG IYRRVAVDYASHTRHVEDIRDTLAETATGISAQAPAYFVSTVSEVVRGAVDGC YTRRLNRQYRFGAATYALLBEGHTVPEVSHNPTVYVPLSLTDTAIGTLRREGGC RRLASMGELFVRGIDVMTAVPAPAGWDLPTVAFERHNTWLEPAEPAASADPLLGT VVSTPGSDRLTAVAAQMSRRAQPMADGILPNAALVEA1RLDGLAGTAPVGLVVDAP VVLPGRGSEVOLIVGEPGQRRRP1EVSRADBPWTRHAGTLPAPAAVPEPAAAG GATDVTYVAGLDADADRYGTHPALDAAVYTTGDDLLDSVMTGVELLASGATVAVTP TATGRLTDPACQPVLYVESVSGTPEVFAOCTGTDLLFDVDPETLDPAPERADELPE ATSAENTISALDAMLADPAETRLAVAVTGDCTEPGAAALMGVLRSAQSEHPRTIVLADL DDPAYLPAVVAAGSEPOVVRNVAASVPRLTPTVPRQDAPLPDPESTVL1GTGTGLGA LTRARHLVAVHGRHLVLSRREAEPLDEELTALGASVA1AACVDAADBAQLEAVLRA1 PAEHP1TAV1HTAGVLDGCVTELTPDRATVVRKPYDAARLDTLEADLAIVLFS SAAGVLTGMPGQAGYAANAELIDALARONSIDLPAVSIANGVYATVSGMETHGSDAD LRRNRIGMSGLPADEGMALDLDA1ATGTTLYAAKFTVDAALRATKAGGPPPLRLGL APUPRAAKTAS1TLERLAGLAETQAAALDLVVRHAEVYGHSGASVSHSGRTFKD AGFDS1TAVELRNLAAATG1LSPAM1FDYKPPALADHLRAKLFGSAAANPAE1GT

KEYWORDS	unidentified.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 53789)
AUTHORS	Schupp,T., Roupet,C. and Engel,N.
TITLE	RIFAMYCIN BIOSYNTHESIS GENE CLUSTER
JOURNAL	Patent: WO 9807868-A 3 26-FEB-1998;
FEATURES	CIBA GEIGY AG (CH)
source	Location/Qualifiers
	1..53789
	/organism="unidentified"
	/db_xref="taxon:32644"
BASE COUNT	6707 a 19183 c 20504 g 7395 t
ORIGIN	
Query Match	47.7%; Score 48.2; DB 6; Length 53789;
Best Local Similarity	67.3%; Pred. No.0.34; Mismatches 33; Indels 0; Gaps 0;
Matches	68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY	1 GTGCAGCCGAGATCGCCGCGCTGCTTCATGAGGGCGCCGAGCGCGCTGCCGTGCAT 60
Db	25279 GTCCGCGCGAGAGTCCGCGCGCTGCTCGGGCACGCCGACGAGCGAGCGCGCTGCAT 25338
OY	61 CGGCGCGCTGTCGACTTGGGCGCTGCATCTCCGCTCAGCGCGGT 101
Db	25339 ACGGCTTCAGAGACGCGCGCTTCGATCGCTGACCGCGGT 25379
RESULT 12	
LOCUS	AF040570 90445 bp DNA Linear BCT 16-APR-2001
AF040570	
DEFINITION	Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.
ACCESSION	AF040570 AF040571
VERSION	AF040570.2 GI:12331604
KEYWORDS	
SOURCE	
ORGANISM	Amycolatopsis mediterranei.
	Amycolatopsis mediterranei
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
	Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
	Amycolatopsis.
REFERENCE	1 (bases 1 to 90445)
AUTHORS	August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R., Yu,T.W.,
	Taylor,M., Hoffman,D., Kim,C.G., Zhang,X., Hutchinson,C.R. and
	Floss,H.G.
TITLE	Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
	the molecular analysis of the rif biosynthetic gene cluster of
	Amycolatopsis mediterranei S699
JOURNAL	Chem. Biol. 5 (2), 69-79 (1998)
MEDLINE	98174059
PUBMED	9512878
REFERENCE	2 (bases 1 to 90445)
AUTHORS	Kim,C.G., Yu,T.W., Fryhle,C.B., Handa,S. and Floss,H.G.
TITLE	3-amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
	formation of the precursor of mcfN units in rifamycin and related
	antibiotics
JOURNAL	J. Biol. Chem. 273 (11), 6030-6040 (1998)
MEDLINE	98165773
PUBMED	9497318
REFERENCE	3 (bases 1 to 90445)
AUTHORS	Yu,T.W., Muller,R., Muller,M., Zhang,X., Draeger,G., Kim,C.G.,
	Leisner,E. and Floss,H.G.
TITLE	Mutational analysis and reconstituted expression of the
	biosynthetic genes involved in the formation of
	3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
	biosynthesis in Amycolatopsis mediterranei S699
JOURNAL	J. Biol. Chem. 276 (16), 12546-12555 (2001)
MEDLINE	21201076
PUBMED	11278540
REFERENCE	4 (bases 1 to 90445)
AUTHORS	August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
	Hutchinson,C.R. and Floss,H.G.
TITLE	Direct Submission

JOURNAL Submitted (30-DEC-1997) Dept. of Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA

REFERENCE 5 (bases 1 to 90445)

AUTHORS August, P. R., Tang, L., Yoon, Y. J., Ning, S., Mueller, R., Hutchinson, C. R. and Floss, H. G.

TITLE Direct Submission

JOURNAL Submitted (19-JAN-2001) Dept. of Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA

REMARK Sequence update by Submitter

REMARK On or before Jan 22, 2001 this sequence version replaced

REMARK gi:2792340, gi:2832263.

FEATURES

SOURCE

1. 90445

/organism="Amycolatopsis mediterranei"

/strain="S699"

/db_xref="taxon:33910"

423. .1451

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="AAC01706.1"

/db_xref="GI:2792310"

/translation="MAPTGVGSGEYVKTLEVALIGAGLILRLHLEAMLAGAIVRVSDGRSRELAEPFGAKAGSILELDCADADITCPASHEITLFTAAAGVAVCEKRLASAEAEIYTAERAGYRLTAANDYRAALYARLHELIVASGRIGEGALGRFSFSYVHPRPWTGHASANSGGILTDQLHAGDLAHVGDVIRHACTQGDATAPBGA AVAGTAVLTHGASGISQVYSRMTATPRPVRAVHVSCTGSGSVSYSEMPQEARVGDGPNFAGVGGPVEFTDMREFATVAFAGGPEPRIGAKDALAAVRIIHAASASAWTGRAV ELPRGAA"

1448. .2215

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="AAC01707.1"

/db_xref="GI:2792311"

/translation="MKVAILSSPTPOAVAGALRGLPDVEVVAAASMDAEPYROAAEAGARVUCETPPRAKETDLMKAMIDAGDRLTFPSPACHGAFAYVRKGTADGIGELITTVLGSVATSDVGVGAAPYLLDLADAVLDGEPQOYVYQOTNTVLSRIGESAAVLTVRRSGVASFDPGRHGSATGLPVTFTIGDGSVQYDAGPOLLGERPELGEDLEALMLMD FLGAGDGPDPGQALRTFRILQAAVESAHGTGVLD"

2353. .2619

/codon_start=1

/transl_table=1

/product="unknown"

/protein_id="AAC01708.1"

/db_xref="GI:2792312"

/translation="MODRSVLGDPDPKOVFAAMSVTOYKIALSGSPALAPLAHRSVDSG ALENTLKIRHGAGCHENFVHSGEFOAVDGCIDIAVFRMSHRTKIAE"

2652. .3842

/note="ORF0; similar to Streptomyces Sp. Chop"

/codon_start=1

/transl_table=1

/product="cytochrome P450 monooxygenase"

/protein_id="AAC01709.1"

/db_xref="GI:2792313"

/translation="MTDAISFEVPMWDTKFDPPAVDSLREERPLAKMYYPDGHWGMLVSSGYELREVLSDLRSHSCSEVGHFVTQGOYIPNPLIPGMFIMDPDEHTRYRKLYSGEYVRAASRLIPRAEVAABQIEVMRAKGAPADVYMDPAFLVLMIGELVGLPYERDRVPAVATLLHDAEADPAEAAAEVAGKFPDEVITERRORPODDLSLSTVEDLTQBELNIVYTLFLFAGYEFTTEGALATGVALLHHTDQLAALAREPKDLDAITELLRLTVNQYHTRYALEDKLEGELIKKQDVTVSLPAA NRDPAKFGCAEALIDEDTSGVHVFGEIHQCLGONLARIELRAGFTALLRAFPETRLAVPADEVPLRLKQSVSVKKL PVSFG"

3883. .18090

/gene="r1fa"

3883. .18090

/gene="r1fa"

/note="r1fa"

/codon_start=1

/transl_table=1

/product="polyketide synthase"

/protein_id="AAC01710.1"

/db_xref="GI:2792314"

gene

CDS

/translaction="NRDILIKPLHVALLENATREPAKPAFAADHRTVYGDLEARTR
LAGHLAGVHRGDRVAICLGNRVSTGAEHVALRAGAVGVLNPGASFALEHPLD
SGATVYVDAADAAARILAPLHVELLVTCDDVPBGASHVDELALSPRAADLEH
PAMFYTSGTIGRKGVYVSTORNCIMSTASYCVPPEPGISDDDRVIMPIPLHLSHIA
CVLSATVYVGAASVRIADSSADVMKRLIEASTSTAGVPTTYHHLVIRARVSGSAPS
LRILAGAVVAGIARLREFEETFGVPLIDAGSTECGATIMNPDPGLRVRSGCSLAP
PGVAVRVDPDPTGDPAGEBEVSGPNMGLGHNSPEATAAMRDMGRTGGLAR
RBDAGYFTFCRIKELIRGGANIHGVEAVLRTVDGADAAVAGVGHDTLGEVLA
YVTPGPGEDPALIEKCRBOLSAVKPDRILTEVAHITPRTASGKIRRLITLDERPQL
YAATEHQSHADESVAAALRARLSGIDERAQCELELDLVRQADVLGQVPEGRA
FRDLEFISLALVELNRNLRTHTGLMPLPASVAFDHPPTPALARAELLGTTQAAEP
VVAADGPEPILAVMACRLPGCVASPEDIMRLVAREVDASEFPDGRMDLSDLPD
REAGTGYVGGGFLHDAGEFDAGFFGSPREAAVMDQOQLLETSMWELNAGVLD
IALKDTGVFSGLMAGOGYSGAVAPLELEGVTGVSASGRSVYGLGEGPAAVY
DTACSSIVAMHILAAORGECSMALAGCTVATBSPFESQORLADGCRKAT
AAADGTGMSGCVYVLERLSVAREKGRILAVIRGSANODGSNLRILPNGISQ
KVRIRALAAAGLAPSDVDVVAHGTGTLSPLEAQLATYGGERQPLMGLSKSN
IGHAQAAGVAVIKMVALNHETLPLTHVADKPTLEVMASAGALIELTEARAMPNG
RPRRAGVSFSGTNAHLIEEAPAEVPAPELPPVPLVVASRSTESLGOERLA
SLBEGDVSLTEVAGALVRRAVLDERAVVAGREAVTGRLALNTAGSGTPKVVY
PFGGTGOWAGMGRLEIAPAEIPARICAAALAPIDMSIVDYRREGDGRVDLO
PACFAYMGLAAWVESVGRDPAVYHSGQETIAACVSAGALSLEDAVVALRSQALA
AELSRGMSAVNLGEDDVSRVLDGVEVAANVGSPSSVALSHDAHALDATLEILSSEG
IYRRVAVDVSHTRHVEDIRDTLAETLAGISAQAPAVPSTVTSIHWBDAGVLDG
IWRNLRNOVFSHTAAATPALLEQGTVEVSAHPYVQPLSEITGDAIGTLRREDGGL
RLSLMGELEFVRGIDVMTAMVPAAGVNDLPTVAFEHRHWLPEAPASAGDPLLGT
VVTSPGSDRLTAIVOMSRRAQPMAYDGLVPAALVEEAIRGLDLAGPVPCEIYVDAP
VVLPRRGSREYQIYGERGEORRRPIEYFSBDEPRTAHAGTILAPAAVPEPAA
GDATDVIVAGLMDADRIGIPHALDAAARIVAGDOLLPESVGLASGATVAVTV
TATGILRTDPAGOPVLIVESVGRPEVAGTETLAPFVDEPILPAPAEADAPJE
ATSAEATLSALOAMLADPAETRLAVNYDCTCEPAGAAIMGVRSSQSHSPRIYIAD
DDPAVLPAVVASGEPOVRNGVAVSPRLPVTGPPRODAPRDEPCTVLTGSGTIGA
LTAHIVYAHGVRHLVYSRRGAPAEIOEELTALGASVATACVADADAEVYRAI
PAERPLAVITHTAGVLDGVYTELPDRLAVRRKDYADALDELTEALDAFVLA
SSAGVLGNPQAGTAAANAEIDLALARKORNLDPAYSIAWVAWVSGMTHEHGDAD
LRNROIGMSGLPADDEGMALDAAIATGCTVAAKFVDAALPATAKAGGAPPLRLG
APLRRAAKASTASTLERLAGLAETEOAALADILVRAHAETLGHSGASVHSRFXD
AGFDSLTAVELRNRLAAATGTLSPAMIEDVPRPALADHRAKIFGANDRPAHEPT
AAAEPIATVAMACRPGCVSPEDIMRLVADGADAVPEPADRQWPDRLXYHDPH
EGTYYVRGATLDADAGDPAFEGISPEBALMDPOOALLIETSELEPRAIDITTL
AGDQIYFAGVANSNDYSMRNRAAGVEBEFRLTGSASVLSGRVAFHVEBPAYVDT
ACSSILVALHMAVOALORGECSMALAGVMVAGVEFESRQGLAPDRCRAFAD
GADGTGMSGEVGLIVERLSEORGRHOVLVAVRGSANVSGASGLTAPASPOQR
IRKALAAAGLSTSDVDAEHAHTGTGTGDPLEAEKELATYCONRPTPLMGSVSNIG
HTOAAAGVAYIKVMAMRHGVLPRTLHVDPSSVYVMSAAVELITARBMVSNRHP
RRAGVSSFGIGTNAHVLEVAAPITTPQEPPEFVLIVLSATYAGLKGQGRILA
AFLEDRTDVRVDAVALATTRAQDLHRAVVLASDRAQLADLAAGSVYGTGVPDG
KLAVLFTGOGSOMAGMGRLEAETEPVEFADAEACAEVDTLREHLEVEFDSALL
DOIVYTGALPAVEATLERLEFESMGVRGLAGHSIGELAAHSGVLDLADAGELVA
ARGHMOALPAGGAMVAVOATEDEVYAPLIDCTVCVAANGDSVYLSCTEAAVAVAD
ELAGRGKTRRLAYSHAHSPLMERMLDDPAVAVRLLTYRAGSLPYVSTINGELAAD
SPDIWGOVRNAVRFSDAVVALGAOGASTLELDGGLAAMALGVLGTPQOSCATL
RKNGAEPDVLTJALAEHVGVDVMTVDEPATAVGTVPJTAFOQRHWFSDIAT
AAVSVPPEPAPYIDRVODVLELVERESAVALGHRDSDLDLBSKSLNLTGTOAAV
KLNRRLBDETVELPSTLIPDYPNPAVLADHLEALDEGERAAPVAPVADTSDPEIAT
VAGMSTRLPBGADSPBELMKLIVABEGDAVSGPVRGMODGLHYHDPDPAHRTASTRSG
GELUDAOFOAGLFGISPREALANDPOORILLIETSELEKABACVDPISARSDBVET
GIVHDIYVTLREVPEDYQITMTGTASVASGRAVYFEGEPRAVYDITACSSILVA
MHILAAALRGECSMALAGATVMASPDALEFSKORLSADGRKAAABEADAGLVA
GLASDVVLEHSGHTGALDPIEQOALLATYGGERQPLMGLSKSLNLTGTOAAV
GVGIMIMAMHGMVAPATLHYDERTSOVMSAGALIVTLTEAREMRPRGRPRACVSP
GASCTNAHLITIEBGPABEAVDEEVAIVPLVYSARASGLAGORLAVLYENESLAG
VAGLVSGRATLNERAVYIAGSRDOADGLOALRGENAGVGTGTAKPCKVWVVP
GQSGOMMGMDLDSFVFAARIKECAALEMOGDSMLDLVLRDADLDRVDVOP
ASPMWVGLAAVMTSLAVTPDAVLGSHGSEITAAACVSGALSILDAKAKVALRQA
ELARGGMAVYALSEEDAVARLTPMANVVEAAVNSPSSVIAGDPAOLDALAEVLA
DGVRVRAVYVYASHTRHVEAIAETLAKTAGIDARVPAIPFYSVIGTWTLEOVVDA
GYMYRNLROQVRGSPVADLAGIHTVVEISAPHVLPQPSLESDAVYGSIRBD
GGLRLLASAEELYRGVAVDMTAVAPPAAGVNDLPTVAFEHRHWLPEAPAEAGM
DGEFWTALIEOSDVSLAELELVEPQORALSTVVPALQMRDRRERHSTAEKLEAEGM
WQPLEREAAGVPGGRMLAVVAGTTDALKEITGCGDILVRLIEHSAIRLADLELRN
VLAEHDLTGLVSLTALDGGPADAAETASTLALVQALDOTTTSATVLMCLTSGAVNIGI

misc_feature		Query Match		47.7%; Score 48.2; DB 1; Length 90445;							
		Best Local Similarity		67.3%; Pred. No. 0.31;	Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;						
QY	1	GTGCAGGCGCGACATGCGCGCGCTGCTTCATGAGGGCGCGGAGCGCGCGCTGCAT	60	Db 27826 GTCCGCGCCAGCTGCGCGCGCTGCGCGCGCACCGCGAGCGCGCTCGCGTCGAC 27885							
Db	27826	GTCCGCGCCAGCTGCGCGCGCTGCGCGCGCTGCGCGCGCACCGCGAGCGCGCTCGCGTCGAC	27885								
QY	61	CGCGCGCTGCTCGACTTGGGCTCGACTGCTCGACGCGCGCGGT	101								
Db	27886	ACGGCCTTCAGAGACGCCGCGCTTCGACCTCGCAGCGCGGT	27926								
RESULT 13											
AF453501/c											
LOCUS											
Actinosynnema pretiosum subsp. auranticum linear BCT 18-JUN-2002											
DEFINITION Actinosynnema pretiosum subsp. auranticum maytansinoid antitumor agent ansimiltoxin biosynthetic gene cluster I, partial sequence.											
ACCESSION AF453501											
VERSION AF453501.1 GI:21449342											
KEYWORDS											
SOURCE											
ORGANISM											
Actinosynnema pretiosum subsp. auranticum.											
Actinosynnema pretiosum subsp. auranticum											
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;											
Actinomycetales; Pseudonocardineae; Actinosynnemataceae;											
Actinosynnema											
REFERENCE											
AUTHORS											
TITLE											
JOURNAL											
MEDLINE											
PUBMED											
REFERENCE											
AUTHORS											
TITLE											
JOURNAL											
FEATURES											
source											
1..82746											
/organism="Actinosynnema pretiosum subsp. auranticum"											
/strain="ATCC 31565"											
/sub_species="auranticum"											
/db_xref="ATCC:31565"											
/db_xref="taxon:42198"											
misc_feature											
1..82653											
/note="cluster I of the maytansinoid antitumor agent ansimiltoxin biosynthetic gene cluster"											
1..1678											
/gene="asm37"											
<1..1678											
/gene="asm37"											
/codon_start=2											
/transl_table=11											
/product="unknown"											
/protein_id="AAM54115.1"											
/db_xref="GI:21449383"											
/translation="IVPGLVVALATVGLAAPASAAITSPVTSTATTVPFVGRTG"											

PAFSRVYVDTDRSAATGYAOGGIGADHLLENGSLYRHSGGMAWTOVRTVPTTSAGV
ARWTVERRADIEQOASPNDDVYQVEAPVETSAKVHTYTPAGAGSYAOPSSANPN
PERGLVHTHGDCADKADLSADALRRYREBOGITLAKMCFYVLAERSSPIATOTALDROR
OPDAARVACVKKILFRFAYTASSESGDAPKREYLAHIDQALILRANADYTAVOAQSF
GAMGENTYTKRNFAGNVSATDQANKRAVQKLLITLPLADRAVQTLPRKRLTIGGG
STRIIGHNDCFLASDTPFGTYEDPATEYPLEAETRLPMGEGTCAVNPRTCPAT
AELSRPHWYLANTDYQODVLRASAGCCLAETVAKLGARTRLDSDVEPTPTTGALIP
RVNLTNDGYAAVNORTVNLVRNTTGTALTRPLTADPRTWLPGARAIADQAQIP
SLATGVYALLLELADPALPQREYSIQTANTGLMDARTGLNDKQITTVS"
1825..2142
/gene="asm36"
/gene="asm36"
1825..2142
/gene="asm36"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54114.1"
/db_xref="GI:21449382"
/translation="MYTASVLCARAGCGAPRPAQATARMFCSTACQOAYRLRKA
GSASHHEDEQLREALLAVRARLTLEATPPWSAQATALVGLARRVVELAARDAGV
SGG"
complement(2129..3307)
/gene="asm35"
complement(2129..3307)
/gene="asm35"
/codon_start=1
/transl_table=11
/product="bicyclomycin resistant protein"
/protein_id="AAM54113.1"
/db_xref="GI:21449381"
/translation="MAVNGADSAVGMPLALLIVALTGTLLIETDMSVSPSPDITARELG
VGGPVALQVTLNFLGCGALGCLSDRFGRRPYLGGSAVYLVALGALCALAPITG
LLTARFVOGIGASAAVLYEVIIGDRYGPALRMYGVALNAAMSLMPCALPGGAI
NRAGRGNYLACATTAJLSILVLEIPEPTRARTAITARSVLADYRSLGSRVLT
AASITPSLLPAALVFIASPRLYTAAGGIVGAPALHLVYASAVBSAFPTGITP
LUGGPAGVARMGAGAVVGYLGLGDDGPAVATGSALECVGFACTVPPDRMSV
FPELAGAASSATGRLALLVTLTLAGALSTGGPPAPAAVMAAVALVLSLSPR
"
complement(3450..4064)
/gene="asm34"
complement(3450..4064)
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54112.1"
/db_xref="GI:21449380"
/translation="MARNRPERAALLDAAEVYLAEGARGLTFRVNDQACVYPACTA
SNFAGRAELITGAGERVYQRLDQAPLDETPUGPRDRSKYTELMHALVDVSAFPPTG
FLALIELRLAARNPELREYLTARADILRFNAEHHAASQLPGDCTTVALMLMML
ILERLTLPDLTDHQRDLVTALVDRLLAGQPEAGSPDESCK"
4151..4723
/gene="asm33"
4151..4723
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AAM54111.1"
/db_xref="GI:21449379"
/translation="MRKLVYFIASIDGVYIAGEGHEDFYPTAEPVEHLRAEYPERTT
PPIHVRPLVGMPLDTPNRRYDVLMMGRGSYQPALDACVSPYALROYVVSRSIPDADP
AAVALRADPVGLRELKREDDLVMLCGGADLAGQLPEIDELVVTYVYVAGGPTPL
PRAGFAPRAFPRPDVRAFDHGVVTTYRPA"
complement(4854..5057)
/gene="asm32"
complement(4854..5057)
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54110.1"
/db_xref="GI:21449378"

/translation="MWITQONARIPARRSSGNSNARCAVWMPQPRSATRAPGRS
SANADRSARGCDPRTSLNRS"
5415..6461
/gene="asm31"
5415..6461
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AAM54109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRRPVRMWGVTEAFWETDEDIKIATLRDAEA
ANADSTRAYLKAIGRVRLTADQEVELARVEAGLYAEHLIDLPEAAGSRRPELR
RVYDQCARAKDHLLEANLRVWSIAKRYTGKAMLLDYOEGNLGLVRAVEKEDYAK
FKSTATWMIROAITRALDOSRTIRLPHVWAEVYTKIARIERDLHRGREATAE
VAVELDTPPEKVELROHARTPESLDOTITGAEDPSIGDLIDENAVAPLEAVSSALL
SDLRVAVLELTPRENGVYRMKIGFPDGVHLITLDEGVRYGVYTRKIRIESTRAKL
QHPARSRGLEGYD"
complement(6501..9518)
/gene="asm30"
complement(6501..9518)
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome p450"
/protein_id="AAM54108.1"
/db_xref="GI:21449376"
/translation="MYATGTRITDGPRLPLNGNLIDVLTSDLDTPVDPLDRCHREHG
ITALITPAGORVYFASSHETVARMCSDPSGKGAHPLAEVDRPAGDGLTTAKDEBNW
GKXNRLMPAFGEPTAMRDHPALMLDIAEOMLVKWRFEQDHRILDAVDQWTRLLDTIA
LCAPGARFNSFEDRAHPEDVAMVRSILVAGERAERLPGVOFLVGRNDRYDDIATM
NRADQIVARARALPAGEPDDILEMLTACADPVGERLSARVRQOLATFLAGHET
TSGLSFEAFALHAPHEVLKAKDAVDGYLGRVPAFEDLARLDYGVYLRFTLRHP
TAPAFALADEPARELEGHAIGAGEPYLVMLPTLRDPAVRDPPDPPRPERDE
IPCANMPRGHGKRAICGRPALQEPATLYLALVLCQFEDLALDPHRLTIKOTLTKP
DSLTVARPRADPGRATVETVPHQVATHRHGTPLVFTGSGNGSGEGARTIAG
DGAARGMATVAPELDVAVALPASGPNVTVSSYNGAPPDNAHFPVMTLQDGPDLG
VYLYLVGCGNLDMASATYQRPVPTLIDEMAMAAGARRLERGATDARDFGDMEWTEP
LMPILSAEGCEVEGEIGPRRVEESDADGLDLSAVVLENRELVGPDASGRHLE
LRPDGTSYRTGDYLSVLPONHDDVLRARVARGPAAEVVYESSAPGLVPVGRAL
RYDELLTRGVDSAPAGACVVARARCPREPRATELATGATLLELERRPSCAVD
LALALELPAFTRRLYSSAAEBQDAEVALVSYTGVTSGTISRPRDRAVAGIAS
PPESFRPADNTVPVVLIAAGTIGAFIRFLAKARALGSEPGALLTFCRPELDDL
YAEFPALDGMLEVDRAYSRHPDGEVRHQHRRDRRELVYAGARVYLCGDATR
VGPAAVEVYGIETPCGAMLDALRAGGRVATDVE"
9664..10245
/gene="asm29"
9664..10245
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54107.1"
/db_xref="GI:21449375"
/translation="MSAAVRREQLVAALRVKRDGIAAATTBAICAEADMPDHAFY
CFRSKQELTYALTGTGLNTSLDAWPAISIPDIPAVDNTYTKMLAAWTARQADPELV
LGEIYTLAOREPELRELPAHQREYVAKVAHVADRRAEABGADYSVDRALLAEIVSV
LQGVASWFSHRGELAKASLDQFAVMPAALR"
10393..10857
/gene="asm28"
10393..10857
/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54106.1"
/db_xref="GI:21449374"
/translation="MTDTYTRHRTYQONALAVGAVFLVGIAGVPGITTYNCDLEF
AGHSDPALLGAVKVSALHNVLHAFGAAGIAGSKARGARAFILGGGAIYAVLML
YGVYVDHSTANFVPLNADNMLHGLAVGMIGLCAALGRNAHRAARASD"
complement(10912..11937)
/gene="asm27"
complement(10912..11937)

Query Match	Best Local Similarity	41.4%;	Score 41.8;	DB 1;	Length 82746;				
Matches	64;	Conservative	0;	Mismatches	37;	Indels	0;	Gaps	0;
QY	1	GTGCAGCCCGAGATCCCGCGTCTTCATCGGGGCGCGCGAGCCCGCTGCGCAT	60						
		11 111 111 1111 11111 111 111 111111 11 1							
Db	28437	GTACAGGGGACACCGCGACCGTGTCTGGGGACCGCAGCGCTCGCGGTGGAGCGGAA	28378						
OY	61	CGGCGGCTGTGCGACTTGGGCGCTCGACTTCCTCAGCGCGGT	101						
		11 11 11 111 1111 111111111111111 11							
Db	28377	CGCGGCTTCAAGGAGCTGGGCGTGCAGCTCGTCACGGCGGT	28337						
RESULT 14									
AX089464	AX089464	16767 bp	DNA	linear	PAT 21-MAR-2001				
DEFINITION	Sequence	49 from Patent WO0116303.							
ACCESSION	AX089464								
VERSION	AX089464.1	GI:13443725							
KEYWORDS									
SOURCE									
ORGANISM	Saccharopolyspora spinosa.								
	Saccharopolyspora spinosa.								
	Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Pseudonocardineae; Pseudonocardaceae;								
	Saccharopolyspora.								
REFERENCE	1 (bases 1 to 16767)								
AUTHORS	Eberz, G., Moehle, V., Froede, R., Veiten, R. and Salas, J.A.								
TITLE	Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis								
JOURNAL	Patent: WO 0116303-A 49 08-MAR-2001;								
	BAYER AG (DE)								
FEATURES	Location/Qualifiers								
source	1..16767								
	/organism="Saccharopolyspora spinosa"								
	/db_xref="taxon:60894"								
	1..16767								
	/note="ORF22; Polyketidesynthase"								
	/codon_start=1								
	/transl_table=1								
	/protein_id="CAC34895.1"								
	/db_xref="GI:13443726"								
CDS									
	/translation="MANEKLREYLKRYVVELEAHERHLEHROEHPIAIVSMGR								
	YPGVSTPEELMRLVNDGGDIAINFPEDRKNMDELDPDGRAGTISVREGGLRG								
	ADPAGLEGISPREAOAMDPOQRLLLEISWEFERAGIDPRLSGTKTVGALVLYH								
	YASRFKRTPAFEEGYFTGNAGSVASGRTVATFGLGPAVVDACSSSLVALAAC								
	SLIRGECDIALAGGISVATPTPGAFVRSRCALASDGRCKFPADADYGTGEGAGMI								
	LLERSDARNGRHVLAAVVGSAINODGTSGLAPSGAPQORYIROLANAGSPAEH								
	VDVNEHSTGALDPTLEAQLIATYGNRSADHPLIGSKSNIGHQAAAGVAV								
	KSVLAIRREMRPSLHIDPSQHDMASGAYRLTDSYDMDLGRPRAGVSFGMSG								
	TNAHLYEVSDEFGVSGTEPTGAPWLSLKTETALPNEQAELLSVTEHPFGLG								
	GVYSIATGRAMERHAAVVADRDSPFAVGLAALAAVAAVGGADCKKVAFFEE								
	GQSGHMMVSLAEIWMRSQGVPAVVGSGEITAAHVAAGLSLTDPAIIAARDCA								
	QLAFFAMVYSLAEIWMRSQGVPAVVGSGEITAAHVAAGLSLTDPAIIAARDCA								
	SALNGKGMIALIPESAVVYRIGRLPGLPTAAVANGSGSTVSGPRLAERLOEIR								
	ENVOTRRKQIDYASHSPIDAOYGRDLRDLRVSSEPAELAFYSTYGERDPTGR								
	DYVONLRQPYRFOOTIYARMADQGRFEVSPHPLTAGIOETLEADAGGVGSD								
	RRGGGSRMRVLSLAECQVRLPVNMGDEVLTNGARVPLPTPFORORYLEAEYD								
	AGDGSVGLISAHPDLGAATVADAGETLLTGKLSYTPQMLADHVYGAILLPTGA								
	EVENIIRADVOGCDLIELSLTTPVLPATGAOVOYLAAGVPDGAARSRVHSR								
	DAVODSMTCHATGTILTSDBHODAGCGDGPIMPNDAAVPLDSFRAAERGPDG								
	AFQGLQAAWKRGDEIFAEVGLPTAHREDAGFGIHPALIDLAOLAGAEEDPDG								
	PAAMQVSLKATGALISLRVLPVPGANAVSVFTTDTTGOAVLSTDSILVKQIS								
	QLAFFAMVYSLAEIWMRSQGVPAVVGSGEITAAHVAAGLSLTDPAIIAARDCA								
	PAARMHESLIRFVPMKRIISPGAAPVSNVAVGNDELNACGSALGTELPDLTGA								

[illegible]

ITAMGGYFRMTEWTPPIGAPLTLEVRTEDCVADEPGRPWTDSDMRPGWTLADATVQV
 PGDHSMDHEHAGSTAQAAVASWLDKLNQRTARQR"
 BASE COUNT 2256 a 4805 c 6457 g 3249 t
 ORIGIN

Query Match 40.0%; Score 40.4; DB 6; Length 16767;
 Best Local Similarity 63.3%; Pred. No. 18;
 Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGGGGCCGCGGAGCGCGCGCGCGCTGCGAT 60
 ||| ||||| ||||| ||| ||| ||||| ||| |||||
 Db 15661 GTCCGCGCCGATATCCGCGGTGACTGGGCGACGCGAGTCGAGCCATGTGACATCGAG 15720

OY 61 CGGCGCGCTGTCGAGCTTGGGCTCGACTCGCTCAGCGC 98
 ||| ||| ||| ||||| ||||| ||||| |||||
 Db 15721 AAGCCTTTGGCGGAGCTGGGTTTCGACTCGCTCAGCGC 15758

RESULT 15
 AX089418

LOCUS AX089418 25360 bp DNA linear PAT 21-MAR-2001
 DEFINITION Sequence 3 from Patent WO0116303.
 ACCESSION AX089418
 VERSION AX089418.1 GI:13443679

KEYWORDS
 SOURCE Saccharopolyspora spinosa.
 ORGANISM Saccharopolyspora spinosa.

REFERENCE 1 (bases 1 to 25360)
 Eberz,G., Moehrle,V., Froede,R., Velten,R. and Salas,J.A.
 Nucleic acids which code for the enzyme activities of the spinosyn
 biosynthesis
 JOURNAL Patent: WO 0116303-A 3 08-MAR-2001;
 BAYER AG (DE)

FEATURES
 source location/Qualifiers
 1..25360
 /organism="Saccharopolyspora spinosa"
 /db_xref="taxon:60894"

BASE COUNT 3832 a 9143 c 8354 g 4031 t
 ORIGIN

Query Match 40.0%; Score 40.4; DB 6; Length 25360;
 Best Local Similarity 63.3%; Pred. No. 17;
 Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 1 GTGCAGCCGAGATCGCGCGCTGCTTTCATGGGGCCGCGGAGCGCGCGCTGCGCGAT 60
 ||| ||||| ||||| ||| ||| ||||| ||| |||||
 Db 92 GTCCGCGCCGATATCCGCGGTGACTGGGCGACGCGAGTCGAGCCATGTGACATCGAG 151

OY 61 CGGCGCGCTGTCGAGCTTGGGCTCGACTCGCTCAGCGC 98
 ||| ||| ||| ||||| ||||| ||||| |||||
 Db 152 AAGCCTTTGGCGGAGCTGGGTTTCGACTCGCTCAGCGC 189

Search completed: November 6, 2002, 13:17:56
 Job time : 618.625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 ; Search time 1197.62 seconds
(without alignments)
1365.823 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cggcggagtcgcccgtacag.....gcttcgcggtgctgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.2	33.9	786	9	AL563086
2	31.4	31.1	1158	13	BM465130
3	31.2	30.9	500	9	AA464685
4	30.8	30.5	910	17	CNS03587
5	30.4	30.1	673	12	BG810602
6	30.2	29.9	819	17	A0840535

C	7	30	29.7	423	14	W79688
	8	30	29.7	427	13	Bj235954
	9	30	29.7	457	13	B1468728
	10	30	29.7	641	13	B1062639
	11	30	29.7	924	12	BG681148
	12	29.8	29.5	611	10	AW972109
	13	29.8	29.5	935	13	B1100110
	14	29.6	29.3	327	13	B1721001
	15	29.6	29.3	378	10	AW415295
	16	29.6	29.3	967	14	BA429819
	17	29.6	29.3	1059	12	BE730429
	18	29.6	29.3	1059	12	BE730429
	19	29.6	29.3	1059	12	BE730429
	20	29.6	29.3	1059	12	BE730429
	21	29.6	29.3	1059	12	BE730429
	22	29.6	29.3	1059	12	BE730429
	23	29.6	29.3	1059	12	BE730429
	24	29.6	29.3	1059	12	BE730429
	25	29.6	29.3	1059	12	BE730429
	26	29.6	29.3	1059	12	BE730429
	27	29.6	29.3	1059	12	BE730429
	28	29.6	29.3	1059	12	BE730429
	29	29.6	29.3	1059	12	BE730429
	30	29.6	29.3	1059	12	BE730429
	31	29.6	29.3	1059	12	BE730429
	32	29.6	29.3	1059	12	BE730429
	33	29.6	29.3	1059	12	BE730429
	34	29.6	29.3	1059	12	BE730429
	35	29.6	29.3	1059	12	BE730429
	36	29.6	29.3	1059	12	BE730429
	37	29.6	29.3	1059	12	BE730429
	38	29.6	29.3	1059	12	BE730429
	39	29.6	29.3	1059	12	BE730429
	40	29.6	29.3	1059	12	BE730429
	41	29.6	29.3	1059	12	BE730429
	42	29.6	29.3	1059	12	BE730429
	43	29.6	29.3	1059	12	BE730429
	44	29.6	29.3	1059	12	BE730429
	45	29.6	29.3	1059	12	BE730429

ALIGNMENTS

RESULT 1
LOCUS AL563086 786 bp mRNA linear EST 16-FEB-2001
DEFINITION AL563086 LTI_NFL003_NBC3 Homo sapiens CDNA clone CSODC025YL03 3
ACCESSION AL563086
VERSION AL563086.1 GI:12912151
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 786)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length CDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
Location/Qualifiers
1..786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODC025YL03"
/clone_id="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

/note="Organ: brain; Vector: pcwvSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcwvSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: filiang@life.com URL: http://fulllength.invitrogen.com"

BASE COUNT 158 a 221 c 216 g 177 t 14 others

ORIGIN

Query Match 33.9%; Score 34.2; DB 9; Length 786;
Best Local Similarity 63.0%; Pred. No. 12;
Matches 51; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

OY 21 GCGGTATCCATCCAGCGGCGCTGATGATGCTGCTTATGACAGACCACCGAC 80
11
DB 181 GGAGTTATMTTGGCATGTCACGACCTATGCGGGTGCTGTGATGACGAGCGACGTA 240
11
OY 81 CGCTCTCGCGGGGTGCGCA 101
11
DB 241 CTCCTCGCGCGGTGCTGCA 261
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

RESULT 2
BM465130 1158 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOUNT_6427971 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5501855
DEFINITION 5', mRNA sequence.
ACCESSION BM465130
VERSION BM465130.1 GI:18514172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1158)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLM12139 row: a column: 24
High quality sequence start: 115
High quality sequence stop: 671.
Location/Qualifiers
1. 1158

FEATURES
source
1. 1158
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5501855"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

BASE COUNT 243 a 364 c 328 g 190 t 33 others

ORIGIN

Query Match 31.1%; Score 31.4; DB 13; Length 1158;
Best Local Similarity 61.7%; Pred. No. 79;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 21 GCGGTATCCATCCAGCGGCGCTGATGATGCTGCTTATGACAGACCACCGAC 80
11
DB 733 GGAGTTATMTTGGCATGTCACGACCTATGCGGGTGCTGTGATGACGAGCGACGTA 674
11
OY 81 CGCTCTCGCGGGGTGCGCA 101
11
DB 673 CTCCTCGCGCGGTGCTGCA 653
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

RESULT 3
AA464685 500 bp mRNA linear EST 10-JUN-1997
LOCUS zx85f11.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
DEFINITION IMAGE:810573 5', mRNA sequence.
ACCESSION AA464685
VERSION AA464685.1 GI:2189569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 500)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m3 rev2 ET from Amer sham
High quality sequence stop: 458.
Location/Qualifiers
1. 500

FEATURES
source
1. 500
/organism="Homo sapiens"
/db_xref="GDB:6041102"
/db_xref="taxon:9606"
/clone="IMAGE:810573"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACATCGAAGTGGAGCGGCGCGGTTTGTGTGTGTGTGTGTGTGTGT 3']; double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 84 a 188 c 133 g 94 t 1 others

ORIGIN

Query Match 30.9%; Score 31.2; DB 9; Length 500;
Best Local Similarity 66.2%; Pred. No. 73;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 33 GCAGCGGAGCGCTGATGATGCTGTGATGACAGACCACCGCTTTCGGGG 92
11
DB 339 GCAGCGGAGCGCTGCGCCAGTGGTGTGAGCTGCATGACGAGCGCTTGGGGG 280
11
OY 93 GTGTGGGC 100
11
DB 279 GTGTGGGC 272
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

RESULT 4
CNS035R7
LOCUS
DEFINITION
    CNS035R7          910 bp    DNA          linear    GSS 15-MAY-2000
    Tetraodon nigroviridis genome survey sequence. PUC-ori end of clone
    214006 of library G from Tetraodon nigroviridis, genomic survey
    sequence.
ACCESSION
    AL229084.1    GI:7888079
VERSION
    GSS: genome survey sequence.
KEYWORDS
    Tetraodon nigroviridis.
SOURCE
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
        Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
        Tetraodontidae; Tetraodon.
REFERENCE
    1 (bases 1 to 910)
    Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
    Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
    Saurin,W. and Weissenbach,J.
    Human gene number estimate provided by genome wide analysis using
    Tetraodon nigroviridis DNA sequence
    Unpublished
    2 (bases 1 to 910)
    Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
    Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
    Weissenbach,J.
    Characterization and repeat analysis of the compact genome of the
    freshwater pufferfish Tetraodon nigroviridis
    Unpublished
JOURNAL
    REFERENCE
    3 (bases 1 to 910)
    Genoscope.
    TITLE
    Direct Submission
    JOURNAL
    COMMENT
    Submitted (12-APR-2000)
    This sequence is a single read and was generated as part of a large
    scale clone-end sequencing project of the Tetraodon nigroviridis
    genome. For more information, please take a look at
    http://www.genoscope.cns.fr/tetraodon.
FEATURES
    source
        1..910
        /organism="Tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clone="214006"
        /clone_1lb="G"
        /note="Genoscope sequence ID : COAG214BH03SP1-end :
        PUC-ori"
BASE COUNT
    188 a      247 c      274 g      196 t      5 others
ORIGIN
    Query Match          30.5%; Score 30.8; DB 17; Length 910;
    Best Local Similarity 59.5%; Pred. No. 1,1e+02;
    Matches 50; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY 17 ACAGGCGTGCATGCAGCCGAGCGCTGATGATGCTGTGATGAGCAGACCAC 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 ACAAGCCAGAGTCTCTGCTGCGCTCTCATCATCATCTACTCTCTGCGAGAGG 857
QY 77 CGACCGCTTCGCGGGGTGCTGCC 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 858 CGTCTCTCCGCGCTGAGAGCACC 881

```

```

REFERENCE
    1 (bases 1 to 673)
AUTHORS
    Choi,W. and Dean,R.A.
TITLE
    Construction and sequence analysis of an appressorium stage cDNA
    library in the rice blast fungus, Magnaporthe grisea
JOURNAL
    COMMENT
    Unpublished (2001)
    Contact: Ralph A. Dean
    Fungal Genomics Laboratory
    North Carolina State University
    Campus Box 7251, Raleigh, NC 27695, USA
    Tel.: 919-513-0020
    Fax: 919-513-0024
    Email: ralph.dean@ncsu.edu
    Seg primer: T3 primer (AATTAACTTCACTTAAGG).
FEATURES
    source
        1..673
        /organism="Magnaporthe grisea"
        /strain="70-15"
        /db_xref="taxon:148305"
        /clone="mgct006x013f"
        /clone_1lb="Magnaporthe grisea Appressorium Stage cDNA"
        /dev_stage="germinated conidia on appressorium-inductive
        surface"
        /note="Vector: pBluescript SK(+). Vector: Site_1: EcoRI;
        Site_2: XhoI. The appressorium formation-specific cDNA
        library was constructed from conidia germinated for 5-8 hr
        on an inductive surface. The library contains over 55,000
        clones with average insert size of 1.5 kbp."
BASE COUNT
    154 a      154 c      232 g      170 t      117 t
ORIGIN
    Query Match          30.1%; Score 30.4; DB 12; Length 673;
    Best Local Similarity 57.3%; Pred. No. 1,3e+02;
    Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 2 GCGGAGTGGCGGTTCAGAGGGCGTATCCATGACCGGAGCGCTCATGATGCTGCT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 GGTGAAGTGGCGGTTCAGAGGGCGTATCCATGATGCTGCTGATGCTGCTGCTGCT 352
QY 62 TGATGAGCAGACACCGACCGCTCTCGCGGGTCT 97
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 351 TGATGCTCTCGCTGCTGACGCTGCTGCTGCTGCT 316

```

```

RESULT 6
AO840535
LOCUS
DEFINITION
    AO840535          819 bp    DNA          linear    GSS 01-SEP-1999
    nbxb0064B04r CUGI Rice BAC Library Oryza sativa genomic clone
    nbxb0064B04r, DNA sequence.
ACCESSION
    AO840535
VERSION
    AO840535.1    GI:5818588
KEYWORDS
    GSS.
SOURCE
    ORGANISM
        Oryza sativa.
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
        Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
    1 (bases 1 to 819)
    Wing,R.A. and Dean,R.A.
    A BAC End Sequencing Framework to Sequence the Rice Genome
    Unpublished (1998)
    Contact: Wing RA
    Clemson University Genomics Institute
    100 Jordan Hall, Clemson, SC 29634, USA
    Tel.: 864 656 7288
    Fax: 864 656 4293
    Email: rwing@clemson.edu
    Seg primer: GGAACAGCTATGACCATG
    Class: BAC ends
    High quality sequence stop: 194.
FEATURES
    source
        1..819
        /organism="Oryza sativa"

```

BASE COUNT
ORIGIN

174 a 216 c 198 g 231 t

```

/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone_md5="0064B04r"
/clone_id="CUGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10b"
/notes="Vector: pBlOBAclI; Site_1: HindIII; Site_2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."
```

		29.9%	Score 30.2;	DB 17;	length 819;
Query Match		69.5%;	Pred. No. 1.6e+02		
Best Local Similarity		0;	Mismatches 18;	Gaps 0;	
Matches	41;	Conservative			
OY	5	GGAGTGGCCCTTTACGAGCGCGTCATGCACGCCGAGCGCCTCATGATGCTGCTTG	63		
Db	297	GGAGTGGCCCTTTACGAGCGCGCTCTCCGTGCACACACAGCAGATGTCATCTCTTT	355		

RESULT	7
LOCUS	w79688/c
DEFINITION	w79688 423 bp mRNA linear EST 17-OCT-1996
ACCESSION	zdb6cc01.t1 Soares-fetal-heart_Nbhh19w Homo sapiens cDNA clone
VERSION	w79688 IMAGE:347520 5', mRNA sequence.
KEYWORDS	w79688.1 GI:1390097
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 423)
Hallier, J., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Motta, M., Parsons, R., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK

```

FEATURES
  source
    Washington University School of Medicine
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: estewatson.wustl.edu
    This clone is available royalty-free through LENT ; contact the
    IMAGE Consortium (info@image.llnl.gov) for further information.
    Insert length: 989      Std Error: 0.00
    Seq primer: mod.REGA+ET
    High quality sequence stop: 384.
    Location/Qualifiers
      1..423
        /organism="Homo sapiens"
        /db_xref="GDB:1272895"
        /db_xref="taxon:9606"

```

					/clone-"IMAGE:347520"						
					/clone_img-"Soares_fetal_heart_NbHH19w"						
					/sex-"unknown"						
					/dev_stage-"19 weeks"						
					/lab_host-"DH10B (ampicillin resistant)"						
					/note-"Organ: heart; Site: pT7SD (Pharmacia) with a modified polylinker; Site1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dN) primer [5' TGTTACCACATCTGAAGTGCGGAGCCGGCAGCTTTTTTTTTTTTTTTT 3'],"						
					double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldi. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHH19w."						
BASE COUNT	87 a	126 c	111 g	95 t	4 others						
ORIGIN											
Query Match		29.7%	Score 30;	DB 14;	Length 423;						
Best Local Similarity	57.4%;	Pred.	No. 1.5e+02:								
Matches	54;	Conservative	0;	Mismatches	40;	Indels	0;	Gaps	0;		
Oy	8 GTGCCCGTTACAGGCGGTGATCATCAAGCCGGAGCGTCGATGATGGTGCCTTGATGA	67									
Dd	250 GTGGGGGCTTCAAGCTGTGGAGAGAGAAGAGAGAGACCACCATGCGAGGAGGCAAAAGGA	191									
Oy	68 GCAGCACCAACGACCGCGTCTCGCGGGGTCTGGCA	101									
Dd	190 GCAACTCACTGCGTAGTGTACAATGCGATTGCA	157									

RESULT 8
 BJ235954
 LOCUS
 DEFINITION
 accession
 version
 keywords
 source
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 BJ235954 427 bp mRNA linear EST 05-APR-2002
 BJ235954 Y. Ogihara unpublished cDNA library, *Whe* Triticum
 aestivum cDNA clone whe19c05 5', mRNA sequence.
 BJ235954
 BJ235954.1 GI:20052145
 EST.
 bread wheat.
 Triticum aestivum
 Euarysta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
 ; Triticeae; Triticum.
 1 (bases 1 to 427)
 Ogihara,Y. and Murai,K.
 Expressed genes in Triticum aestivum
 Unpublished (2002)
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..427

```

/organism="Prilicium aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="phei9c05"
/clone_lib="Y. Ogihara unpublished cDNA library, Wh_e"
/tissue_type="seed DP10"
/dev_stage="Feekes, scale 11.2"
/notes="Vector: Lambda uni-zap XR, excised phagemid;
Site:1: EcoRI; Site:2: XhoI; Plants were grown under
hydroponic conditions at UC Davis, salt stressed for 12
hours, and for 7 days, then dissected and frozen (Akhnouf
in J. Dvorak Lab). Total RNA was prepared from sheath
tissue, equal quantites of RNA were pooled from the two
samples, polyA was purified from the pooled RNA, a cDNA

```

library was made, and the cDNA clones were in vivo excised to give pluescript phagemids in the T3 Close lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT
ORIGIN

Query Match 29.7%; Score 30; DB 13; Length 427;
Best Local Similarity 61.5%; Pred. NO. 1.5e+02;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 3 CGCGAGTGGCGGCTTACGATGATGACCGGAGCGGATGATGATGATGCT 62
Db 74 GCCGAGTGGCGGCTTACGATGATGATGATGATGATGATGATGATG 133

OY 63 GATGAGCAGACCGAC 80
Db 134 ATGTGACGAGACCGGC 151

RESULT 9
BI468728
LOCUS
DEFINITION
ID: Gm-cl050-4474.5, similar to SW:ATP4_IP0BA Q40089 ATP SYNTHASE
DELTA CHAIN, MITOCHONDRIAL PRECURSOR; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
BI468728.1 GI:15284837

soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE
AUTHORS
1 (bases 1 to 457)
Shoemaker, R., Keim, P., Vodkin, L., Erpelidng, J., Corvett, V., Khanna
'A', Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, D., Beck, C.,
Wyle, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
'Y', Person, B., Swaller, T., Gibbons, M., Pape, D., Harrey, N., Schurk
'R', Ritter, E., Kohn, S., Shiu, T., Jackson, Y., Cardenas, M., McCann
'R', Waterson, R. and Wilson, R.

TITLE
JOURNAL
COMMENT
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccreesgen.com
High quality sequence stop: 425.

FEATURES
SOURCE

Location/Qualifiers
1. 457
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl050-4474"
/clone_lib="Gm-cl050"
/tissue_type="leaf tissue at various developmental stages
of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pluescript II SK+, site 1: EcoRI; site 2:
XhoI; The Clark NIL was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from leaf tissue at various developmental stages
of 3 week old greenhouse grown plants. Complementary DNA

BASE COUNT
ORIGIN
91 a 159 c 106 g 101 t

Query Match 29.7%; Score 30; DB 13; Length 457;
Best Local Similarity 59.3%; Pred. NO. 1.5e+02;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 1 CGCGAGTGGCGGCTTACGATGATGATGACCGGAGCGGCTGATGATGATGCT 60
Db 166 CGGAGCGCTGGAAGAAGTGAAGCTTCTTCATGACCGCGGAGACCGCGCTGCGTAC 225

OY 61 TTGATGAGCAGACGACGACCGCTTC 86
Db 226 ATGAAGCCCCGACACCGACCTCTTC 251

RESULT 10
BI062639
LOCUS
DEFINITION
ID: UT00117-210301-510-C12 UT0117 Homo sapiens CDNA, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
BI062639.1 GI:14470166

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 641)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
'M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE
JOURNAL
COMMENT
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3<=IL3-UT00117-
210301-510-C12&lc=2001-03-21<=1)
Seq primer: puc 18 forward
High quality sequence stop: 514.

FEATURES
SOURCE

Location/Qualifiers
1. 641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT0117"
/dev_stage="Adult"
/note="Organ: uterus, tumor. Vector: puc18; site 1: SmaI;
site 2: SmaI; A mini-library was made by cloning products
derived from ORFESTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of

Db 85 GGAGATTGGCCAGAGAGGTCACAAACAGAGAGTCGAGAGTTGGGTGGTGCCCT 144
QY 62 TGATGAGCAGACCACC 77
||| |||
Db 145 GAGCACACAGAGCACC 160

Search completed: November 6, 2002, 15:52:22
Job time : 1202.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:32:49 : Search time 27.125 Seconds
(without alignments)
1240.503 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cgcgcgagtgccgtacag.....gtcttcgcggtgcgcga 101

Scoring table: IDENTITY,NUC

Gapop 10.0 , Gapept 1.0

Searched: 310279 seqs, 16577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCr_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	37	36.6	13842	10	US-09-861-289-30
2	37	36.6	36778	10	US-09-861-289-5
3	26.8	26.5	263	10	US-09-923-876-3525
4	26.6	26.3	202	10	US-09-815-242-1105
5	26.6	26.3	443	10	US-09-815-242-1258
6	26.6	26.3	540	10	US-09-815-242-4112
7	26.4	26.1	538	10	US-09-771-1614-70
8	26.4	26.1	2457	12	US-10-052-586-569
9	26.4	26.1	13224	10	US-09-764-853-897
10	26	25.7	260	10	US-09-815-242-1239
11	26	25.7	556	10	US-09-873-880-21
12	26	25.7	1254	10	US-09-815-242-6157
13	26	25.7	15872	10	US-09-861-289-1
14	26	25.7	32204	10	US-09-764-855-327
15	25.8	25.5	2968	10	US-09-815-915-1
16	25.6	25.3	412	10	US-09-860-352-12815
17	25.6	25.3	414	10	US-09-860-352-12671
18	25.6	25.3	429	10	US-09-860-352-7004
19	25.6	25.3	1501	10	US-09-768-894A-1

20	25.6	25.3	6503	10	US-09-935-541-12	Sequence 12, Appl
21	25.4	25.1	395	10	US-09-867-701-6689	Sequence 6689, Ap
22	25.4	25.1	419	10	US-09-867-701-6858	Sequence 6858, Ap
23	25.4	25.1	1561	10	US-09-729-674-177	Sequence 177, App
24	25.4	25.1	1677	10	US-09-915-582-16	Sequence 16, Appl
25	25.4	25.1	1802	10	US-09-799-777-104	Sequence 104, App
26	25.4	25.1	3012	10	US-09-767-215-3	Sequence 3, Appl1
27	25.4	25.1	3409	10	US-09-954-456-501	Sequence 501, App
28	25.4	25.1	3417	10	US-09-767-215-4	Sequence 4, Appl1
29	25.4	25.1	3931	10	US-09-767-215-1	Sequence 1, Appl1
30	25.2	25.0	232	10	US-09-864-761-31781	Sequence 31781, A
31	25.2	25.0	512	10	US-09-864-761-15259	Sequence 15259, A
32	25.2	25.0	1738	10	US-09-764-877-3218	Sequence 3218, Ap
33	25	24.8	1848	10	US-09-915-582-34	Sequence 34, Appl
34	24.8	24.6	398	10	US-09-860-352-1749	Sequence 1749, Ap
35	24.6	24.4	531	10	US-09-864-761-6877	Sequence 6877, Ap
36	24.6	24.4	1845	10	US-09-822-849A-292	Sequence 292, App
37	24.4	24.2	263	10	US-09-878-574-15767	Sequence 15767, A
38	24.4	24.2	933	9	US-09-934-392-3	Sequence 3, Appl1
39	24.4	24.2	1435	9	US-09-813-319A-1	Sequence 1, Appl1
40	24.4	24.2	1543	9	US-09-934-392-1	Sequence 1, Appl1
41	24.4	24.2	1566	10	US-09-833-381-987	Sequence 987, App
42	24.4	24.2	1700	10	US-09-817-184-1	Sequence 1, Appl1
43	24.4	24.2	3209	10	US-09-925-301-474	Sequence 474, App
44	24.4	24.2	24881	10	US-09-817-184-3	Sequence 3, Appl1
45	24.4	24.2	29695	10	US-09-752-820A-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861.289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match 36.6%; Score 37; DB 10; Length 13842;
Best Local Similarity 62.4%; Pred. No. 0.0046;
Matches 58; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 CGCGGAGTGGCCCTTACAGGCGGTATCATGCAGCGGAGCGCTGATGATGCTGTC 60
DB 6848 CGCGGAGTGGCCCTTACAGGCGGTATCATGCAGCGGAGCGCTGATGATGCTGTC 6907

QY 61 TTGATGACGACACCGACCGACCGCTTCTGCGGG 93
DB 6908 TGGACACGCTGACCGCGGACGACGATCGCGGG 6940

RESULT 2
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```

; APPLICANT: Liu, H.
; APPLICANT: xue, Y.
; APPLICANT: zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```

Query Match          36.6%; Score 37; DB 10; Length 36778;
Best Local Similarity 62.4%; Pred. No. 0.0054;
Matches 58; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```

QY 1 CGCGGAGTGGCGCTTACAGGCGGTGATCCATGACGCCGGGCGTGGATGATGATGTC 60
DB 8589 CGCGCGAGACGCCCTTCACCGCCGTCTCCACACCGCGCGCTCGACGAGCATCG 8648
QY 61 TTGATGACGACGACCGACCGACCGCTTCGCGCG 93
DB 8649 TGGACACGCTGACCGCGGACGACGATTCGCGCGG 8681
```

```

RESULT 3
US-09-923-876-3525/c
; Sequence 3525, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO: 3525
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700162447H1
US-09-923-876-3525
```

```

Query Match          26.5%; Score 26.8; DB 10; Length 263;
Best Local Similarity 55.3%; Pred. No. 4.5;
Matches 52; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```

QY 5 GGAGTGGCGCTTACAGGCGGTGATCCATGACCGGAGCGCTGATGATGATGTCGTTGA 64
DB 95 GCGGAGCGTGTGAGTAGAGAGCGAGCGTAAAGAGGCGGAGGCGGATTTCTCGCG 36
QY 65 TGACGACGACCGACCGCGCTTCGCGGGGTCTG 98
DB 35 TGGCGGCTTCTGAACCGGAGATTGGGCGGCTG 2
```

```

RESULT 4
US-09-815-242-1105
; Sequence 1105, Application US/09815242
```

```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 1105
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-1105
```

```

Query Match          26.3%; Score 26.6; DB 10; Length 202;
Best Local Similarity 58.0%; Pred. No. 5;
Matches 47; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```

QY 6 GAGTGGCGCTTACAGGCGGTGATCCATGACCGGAGCGCTGATGATGTCGTTGAT 65
DB 68 GAAAGCAGCGCAGCGAGCGGCTTCATCCATCCGACGAGCGGTGTGTCGATGAT 127
QY 66 GAGCAGACCGACCGACCGCTTC 86
DB 128 GTCCAGGCCACGACGCGCATC 148
```

```

RESULT 5
US-09-815-242-1258
; Sequence 1258, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

Query Match	26.3%	Score	26.6	DB	10	Length	540
Best Local Similarity	58.0%	Pred. No.	5.7				
Matches	47	Conservative	0	Mismatches	34	Indels	0
						Gaps	0

; PRIOR FILING DATE: 1997-09-18
 ; PRIOR APPLICATION NUMBER: 60/059266
 ; PRIOR FILING DATE: 1997-09-18
 ; PRIOR APPLICATION NUMBER: 60/062250

PRIOR FILING DATE: 1997-10-17	PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24	PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24	PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21	PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28	PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29	PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31	PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31	PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13	PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21	PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24	PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24	PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11	PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12	PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17	PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18	PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10	PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11	PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11	PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01	PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01	PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08	PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08	PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21	PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21	PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22	
PRIOR APPLICATION NUMBER: 60/083222	
PRIOR APPLICATION NUMBER: 60/083495	
PRIOR APPLICATION NUMBER: 60/083496	
PRIOR APPLICATION NUMBER: 60/083499	
PRIOR APPLICATION NUMBER: 60/083559	
PRIOR APPLICATION NUMBER: 60/084229	
PRIOR APPLICATION NUMBER: 60/084366	
PRIOR APPLICATION NUMBER: 60/084414	
PRIOR APPLICATION NUMBER: 60/084639	
PRIOR APPLICATION NUMBER: 60/084639	
PRIOR APPLICATION NUMBER: 60/084640	
PRIOR APPLICATION NUMBER: 60/084643	
PRIOR APPLICATION NUMBER: 60/085580	
PRIOR APPLICATION NUMBER: 60/085582	
PRIOR APPLICATION NUMBER: 60/085700	
PRIOR APPLICATION NUMBER: 60/086023	
PRIOR APPLICATION NUMBER: 60/086392	
PRIOR APPLICATION NUMBER: 60/086486	
PRIOR APPLICATION NUMBER: 60/087098	
PRIOR APPLICATION NUMBER: 60/087208	
PRIOR APPLICATION NUMBER: 60/087609	
PRIOR APPLICATION NUMBER: 60/087759	
PRIOR APPLICATION NUMBER: 60/087939	
PRIOR APPLICATION NUMBER: 60/087827	
PRIOR APPLICATION NUMBER: 60/088025	
PRIOR APPLICATION NUMBER: 60/088025	
PRIOR APPLICATION NUMBER: 60/088028	
PRIOR APPLICATION NUMBER: 60/088217	
PRIOR APPLICATION NUMBER: 60/088217	
PRIOR APPLICATION NUMBER: 60/088222	
PRIOR APPLICATION NUMBER: 60/088326	
PRIOR APPLICATION NUMBER: 60/088555	
PRIOR APPLICATION NUMBER: 60/088722	
PRIOR APPLICATION NUMBER: 60/088738	
PRIOR APPLICATION NUMBER: 60/088740	


```

: ORGANISM: Trifolium aestivum
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (208)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (330)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (357)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (360)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (362)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (389)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (430)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (437)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (467)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (491)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (502)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (515)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (520)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (548)
: OTHER INFORMATION: n=A, C, G, or T
: NAME/KEY: unsure
: LOCATION: (551)
: OTHER INFORMATION: n=A, C, G, or T
:
: OS-09-873-880-21

```

```

Query Match      25.7%; Score 26; DB 10; Length 556;
Best Local Similarity 53.5%; Pred. NO. 9;
Matches 53; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY      2  GCGCGAGTGGCGGCTTACAGGGCGGCGTATCATGACCCGGAGCGCTCGATGATGATGTGCT 61
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      123 GCGCGCGGAGGCGCATGCGGGCGGCCCATGAGCGCGGCGGAGCGTGGACGACGACGTGCTGGG 182
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      62  TGATGACGACGACACCGACCGCTTCTCCGGGTGCTGGC 100
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      183 GCGCGACCCGACGCGGCGTCCGCTTCNAGCGGAGATGGC 221
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-815-242-6157/c
; Sequence 6157, Application US/09815242
; Patent No. US20020061369A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

```

```

APPLICANT : Yamamoto, Robert T.
APPLICANT : Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6157
LENGTH: 1254
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1254)
US-09-815-242-6157

```

[illegible]

Db 7566 CCGCGGAGCATCGCTGACGCGCCGTCGTGCACACGGCGGAGTGTGCAGCAGCGAGCG3 7625
QY 61 TTGATGAGACAGACCCAGCCGCTTCTGCGGGTGTG 98
Db 7626 TGGAGGCGCTCACACCGGAGCGGCTGTGACGCGTACTG 7663

RESULT 14

US-09-764-855-327/c
; Sequence 327, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 327
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-327

Query Match 25.7%; Score 26; DB 10; Length 32204;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 50; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 GCGAGTGGCGCTTACAGGGGCTGATCATGTCAGCGGAGCGCTCGATGATGTGTGCTT 62
Db 102 GCGAGGCGCGCTGTGCGGGTGGCGTGGGTCCAGAGCGAGCGCGGAGGCGCGGGG3 43
QY 63 GATGAGCAGACCGACCGACCGCTTCTCGCGG 92
Db 42 GCCGCGGAGGCGCCCGGACCGCTTCTGCGCG 13

RESULT 15

US-09-815-915-1/c
; Sequence 1, Application US/09815915
; Patent No. US20020025931A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 3714, 16742, 23546, AND 13887 NOVEL
; FILE REFERENCE: 38155-20006.00
; CURRENT APPLICATION NUMBER: US/09/815,915
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,846
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2968
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)...(2491)
; NAME/KEY: misc_feature
; LOCATION: (1)...(2968)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-915-1

Query Match 25.5%; Score 25.8; DB 10; Length 2968;
Best Local Similarity 56.5%; Pred. No. 13;
Matches 48; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 CGAGTGGCGCTTACAGGGGCTGATCATGTCAGCGGAGCGCGCTCGATGATGTGTGCTT 63

Db 207 CGAGGGGCTTCTGTGCTGGCCCTGACCTTGGCCCGGGGCTCCGGGCTGAACCTCCGACATG 148
QY 64 ATGACAGACCGACCGACCGCTTCTC 88
Db 147 ATAACATGCTCCGCGGCACCTC 123

Search completed: November 6, 2002, 20:30:50
Job time : 51.125 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:54 : Search time 33.5 Seconds
(without alignments)
924.608 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cgcgcgagtgccgttaccag.....gtcttcgcggtgtcgca 101

Scoring table: IDENTITY_NUC

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents.NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCtUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	3 US-09-335-409-1	Sequence 1, Appli
2	101	100.0	68750	4 US-09-568-102-1	Sequence 1, Appli
3	101	100.0	68750	4 US-09-567-969-1	Sequence 1, Appli
4	101	100.0	68750	4 US-09-568-480-1	Sequence 1, Appli
5	101	100.0	68750	4 US-09-568-486-1	Sequence 1, Appli
6	101	100.0	68750	4 US-09-568-472-1	Sequence 1, Appli
7	101	100.0	68750	4 US-09-567-899-1	Sequence 1, Appli
8	101	100.0	71989	4 US-09-443-501A-2	Sequence 2, Appli
9	51.6	51.1	33529	4 US-09-144-085-3	Sequence 3, Appli
10	44	43.6	49377	4 US-08-764-233A-1	Sequence 1, Appli
11	41.2	40.8	43280	2 US-08-804-227C-1	Sequence 1, Appli
12	38.8	38.4	13987	2 US-08-804-227C-13	Sequence 13, Appli
13	38.8	38.4	44377	2 US-08-804-227C-7	Sequence 7, Appli
14	38.8	38.4	44377	2 US-08-804-198-1	Sequence 1, Appli
15	38.8	38.4	80161	3 US-09-036-987A-1	Sequence 1, Appli
16	38.8	38.4	80161	4 US-09-370-700-1	Sequence 1, Appli
17	37	36.6	13842	4 US-09-105-537-30	Sequence 30, Appli
18	37	36.6	36778	4 US-09-105-537-5	Sequence 5, Appli
19	37	36.6	38506	3 US-09-320-878-19	Sequence 19, Appli
20	35	34.7	28958	1 US-08-258-261B-6	Sequence 6, Appli
21	35	34.7	28958	1 US-08-456-837-6	Sequence 6, Appli
22	35	34.7	28958	1 US-08-457-342-6	Sequence 6, Appli
23	35	34.7	28958	1 US-08-457-646A-6	Sequence 6, Appli
24	35	34.7	28958	1 US-08-458-076A-6	Sequence 6, Appli
25	35	34.7	28958	1 US-08-764-233A-4	Sequence 4, Appli
26	35	34.7	28958	1 US-08-457-335A-6	Sequence 6, Appli
27	35	34.7	28958	1 US-08-729-214-6	Sequence 6, Appli

28	35	34.7	28958	3 US-09-028-934-6	Sequence 6, Appli
29	32.4	32.1	50937	4 US-09-428-517-1	Sequence 1, Appli
c	30	32.4	4403765	4 US-09-103-840A-2	Sequence 2, Appli
31	30.8	30.5	11219	1 US-07-642-734C-1	Sequence 1, Appli
32	30.8	30.5	11219	1 US-08-439-009A-1	Sequence 5, Appli
33	28.4	28.1	2493	1 US-07-977-434-5	Sequence 5, Appli
34	28.4	28.1	2493	1 US-08-458-819-5	Sequence 5, Appli
35	28.4	28.1	2493	5 PCT-US91-07035-5	Sequence 5, Appli
36	28.2	27.9	70000	4 US-09-851-896-3	Sequence 3, Appli
37	27.4	27.1	2634	3 US-08-911-853-30	Sequence 30, Appli
38	27.4	27.1	2634	4 US-09-479-409-30	Sequence 30, Appli
39	27.4	27.1	2634	4 US-09-479-453-30	Sequence 29, Appli
40	27.4	27.1	17612	3 US-08-911-853-29	Sequence 29, Appli
41	27.4	27.1	17612	4 US-09-479-409-29	Sequence 29, Appli
42	27.4	27.1	17612	4 US-09-479-453-29	Sequence 29, Appli
c	43	26.8	1811	4 US-09-280-116-94	Sequence 94, Appli
c	44	26.8	2085	3 US-09-299-843A-65	Sequence 65, Appli
c	45	26.8	2085	4 US-09-088-337B-65	Sequence 65, Appli

ALIGNMENTS

RESULT 1
US-09-335-409-1
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335.409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1
Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 CGCGGAGTGGCGCTTACAGGGCGTGATCATGACAGCCGAGCGCTCATGATGCTGTGC 60
Db 36162 CGCGGAGTGGCGCTTACAGGGCGTGATCATGACAGCCGAGCGCTCATGATGCTGTGC 36221
Oy 61 TTGATGACAGACACGACGCGCTTCTCGCGGGTCTGTGCA 101
Db 36222 TTGATGACAGACACGACGCGCTTCTCGCGGGTCTGTGCA 36262
RESULT 2
US-09-568-102-1
Sequence 1, Application US/09568102
Patent No. 6346404
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568.102

```
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 36262
```

```
RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 36262
```

```
RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
```

```
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 36262
```

```
RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.7e-23;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 60
          |||||||
Db 36162 CGGCGGAGTGCCGCTTACAGGGCGGTGATCGACGGGAGCGCTCGATGATGTTGTC 36221
```

```
OY      61 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 101
          |||||||
Db 36222 TTGATGAGCAGACCAACCGACCGCTTCTCGGGGTGCTGGCA 36262
```

```
RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
```

: APPLICANT: Zirkle, Ross
 : APPLICANT: Cyr, Devon
 : APPLICANT: Goerlach, Joern
 : TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 : FILE REFERENCE: 4-30582A
 : CURRENT APPLICATION NUMBER: 09/335,409
 : CURRENT FILING DATE: 2000-05-10
 : PRIOR APPLICATION NUMBER: 09/335,409
 : PRIOR FILING DATE: 1999-06-17
 : NUMBER OF SEQ ID NOS: 30
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 1
 : LENGTH: 68750
 : TYPE: DNA
 : ORGANISM: Sorangium cellulosum
 : US-09-568-472-1

Query Match	100.0%	Score 101;	DB 4;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 2.7e-23;		
Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 CGCGGAGTGGCCGTTACAGGGCGTGATCATGCACCGGAGCGCTCGATGATGGTGTGC 60
 |||||
 Db 36162 CGCGGAGTGGCCGTTACAGGGCGTGATCATGCACCGGAGCGCTCGATGATGGTGTGC 360

61 TTGATGAGCAGACCACCGCTTCTCGCGGTGCTGCA 101

Db 36222 TTGATGAGCAGACCACCGCGTCTCTCGGGGTGCTGGCA 36262

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899

```

? GENERAL INFORMATION:
? APPLICANT: Schupp, Thomas
? APPLICANT: Liigon, James
? APPLICANT: Molnar, Istvan
? APPLICANT: Zirkle, Ross
? APPLICANT: Cyr, Devon
? APPLICANT: Goerlach, Joern
? TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
? FILE REFERENCE: 4-30562A
? CURRENT APPLICATION NUMBER: US/09/567,899
? PRIOR FILING DATE: 2000-05-10
? PRIOR APPLICATION NUMBER: 09/335,409
? PRIOR FILING DATE: 1999-06-17
? NUMBER OF SEQ ID NOS: 30
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 68750
? TYPE: DNA
? ORGANISM: Sorangium cellulosum
? US-09-567-899-1

```

Query March	100.0%	Score 101;	DB 4;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 2.7e-23;		
Matches 101; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy 1 CGGGCGAGTGGCCGTTACAGGCGGTGATCCATGCAGCCGGAGCGCTGCATGATGGTGTGC 60
|||||
Db 36162 CGGGCGAGTGGCCGTTACAGGCGGTGATCCATGCAGCCGGAGCGCTGCATGATGGTGTGC 36222

61 TTGATGAGCAGACCACCGACTTCTCGGGGTGCTGGCA 101

Db 36222 TTGATGAGCAGACCAACCGCTTCTCGCGGGTGCTGGCA 36262

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:

```

APPLICANT: Kosan Biosciences, Inc.
APPLICANT: Julien, Bryan
APPLICANT: Katz, Leonard
APPLICANT: Khosla, Chaitan
APPLICANT: Tang, Li
APPLICANT: Ziemann, Rainer
TITLE OF INVENTION: Recombinant Methods and Materials for Producing
TITLE OF INVENTION: Epothilone Derivatives
FILE REFERENCE: 30062-20031.00
CURRENT APPLICATION NUMBER: US/09/443,501A
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

```

Query Match	100.0%	Score 101;	DB 4;	Length 71989;
Best Local Similarity	100.0%	Pred. No. 2.7e-23;		
Matches 101; Conservative	0;	Mismatches	0;	Gaps 0;

Qy 1 CGCGGAGTGGCCGTTACAGGCGGTGATCCATGACGCCGAGCCCTCGATGATGGTGTGC 60
|||||
|||
Db 30550 CGCGGAGTGGCCGTTACAGGCGGTGATCCATGACGCCGAGCCCTCGATGATGGTGTGC 30609

61 TTGATGAGCAGACCACCGCTTCTCGGGGTGTCGCA 101

Db 30610 TTGATGAGCAGACCACCGCTTCTCGCGGTGCTGCA 30650

RESULT 9
US-09-144-085-3
; Sequence 3, Application US/09144085

```

: GENERAL INFORMATION:
: APPLICANT: Gustafsson, Claes
: APPLICANT: Betlach, Mary C.
: APPLICANT: Ashley, Gary
: APPLICANT: Julien, Bryan
: APPLICANT: Ziemann, Rainer
: TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
: TITLE OF INVENTION: THREEFOR
: FILE REFERENCE: 30062.20020.20
: CURRENT APPLICATION NUMBER: 05/09/144.085
: CURRENT FILING DATE: 1998-08-31
: EARLIER APPLICATION NUMBER: 09/010.809
: EARLIER FILING DATE: 1998-01-22
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 33529
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: OS-09-144-085-3

```

	Query Match	51.1%	Score 51.6	DB 4	Length 33529
	Best Local Similarity	70.4%	Pred. No. 1.5e-07		
	Matches 69	Conservative 0	Mismatches 29	Indels 0	Gaps 0
0y	3 GCGAGTGGCCGTTACAGGCGCTATCCATGACGCCGAGCCCTCGATGATGTGTGCTT	62			


```

: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 43280 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 816..14234
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14351..19945
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20010..31199
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31232..36067
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36249..41774
:
: US-08-804-227C-1
:
: Query Match 40.8%; Score 41.2; DB 2; Length 43280;
: Best Local Similarity 67.4%; Pred. No. 0.00035;
: Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
:
: QY 1 CGGGGAGTGGCGTTACAGGCGGTGATCGACGCCGAGCGCTCGATGATGTGTGC 60
: Db 13093 CGGGCGACGGCGCGTGCAGCGAGTGCACGCGCGCGGCGTCTCGACAGCGGCACAC 13152
:
: QY 61 TTGATGACAGACACCGACCGCTTC 86
: Db 13153 TCGACGGCGCTGACCCCGGAGACGATC 13178
:
: RESULT 12
: US-08-804-227C-13
: Sequence 13, Application US/08804227C
: Patent No. 5876991
:
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS

```

```

: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 13987 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..13987
:
: US-08-804-227C-13
:
: Query Match 38.4%; Score 38.8; DB 2; Length 13987;
: Best Local Similarity 62.2%; Pred. No. 0.0016;
: Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
:
: QY 1 CGGGGAGTGGCGCGTTACAGGCGGTGATCGACGCCGAGCGCTCGATGATGTGTGC 60
: Db 7260 CCGAGAGCGGCGCGCTGCTCGCGTACTGCACGCGGCGAGTGTCTGACGACGCGTGTGC 7319
:
: QY 61 TTGATGACAGACACCGACCGCTTCGCGGCTTC 98
: Db 7320 TCGACTCGCTGACCTCGACGCGGCTGACGCGCTGACTG 7357
:
: RESULT 13
: US-08-804-227C-7
: Sequence 7, Application US/08804227C
: Patent No. 5876991
:
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784

```

REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match Best Local Similarity 38.4%; Score 38.8; DB 2; Length 44377;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGAGCGGTATCCATCCAGCCGAGCGCTCGATGATGTGTGC 60
DB 7275 CCGAGGAGGGGCCCTCGTCCGCTACTGACGCGGAGGTGTGTCGAGAGGGTGTGC 7334
QY 61 TTGATGACGACGACCGACCGCGCTTCTCGGGGTGCTG 98
DB 7335 TCGACTGCTCACCTCCGACCGGGTGGAGCCCGCTACTG 7372

RESULT 14

US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostek, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match Best Local Similarity 38.4%; Score 38.8; DB 2; Length 44377;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGAGCGGTATCCATCCAGCCGAGCGCTCGATGATGTGTGC 60
DB 7275 CCGAGGAGGGGCCCTCGTCCGCTACTGACGCGGAGGTGTGTCGAGAGGGTGTGC 7334
QY 61 TTGATGACGACGACCGACCGCGCTTCTCGGGGTGCTG 98
DB 7335 TCGACTGCTCACCTCCGACCGGGTGGAGCCCGCTACTG 7372

RESULT 15

US-09-036-987A-1
Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madhuri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479


```

; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

```

```

Query Match 38.4%; Score 38.8; DB 3; Length 80161;
Best Local Similarity 62.2%; Pred. No. 0.0023;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```

```

QY 1 CGGCGAGTGGCGTTACAGGGCGTGATCATGACGCGAGCGCTCGATGATGGTGTGC 60
    ||||| ||||| || || || || || || || || || || || || || || || ||
Db 74828 CTGGGGAACCGCTGTGGGGTGGTGTGCACGCCGCCGGTGTGTGATGACGGTGTGC 74887
    ||||| ||||| || || || || || || || || || || || || || || || ||
QY 61 TTGATGAGCAGACCAACCGCGCTTCTCGCGGGTGTG 98
    || || || || || || || || || || || || || || || || || || || ||
Db 74888 TGATGTGATGTGCGCGGAGCGCTTGACGCGGTGTG 74925
    || || || || || || || || || || || || || || || || || || || ||

```

Search completed: November 6, 2002, 15:59:48
Job time : 201.625 secs

FT	/product= "Orf 5 protein (AAV58584) "
FT	CDS
FT	/tag- f
FT	/product= "Type I polyketide synthase, EPOS A
FT	(AAV58573) "
FT	CDS
FT	11872..116104
FT	/tag- g
FT	/product= "Non-ribosomal peptide synthetase, EPOS P
FT	(AAV58574) "
FT	CDS
FT	16251..21749
FT	/tag- h
FT	/product= "Type I polyketide synthase, EPOS B
FT	(AAV58575) "
FT	CDS
FT	21746..43519
FT	/tag- i
FT	/product= "Type I polyketide synthase, EPOS C
FT	(AAV58576) "
FT	CDS
FT	43524..54920
FT	/tag- j
FT	/product= "Type I polyketide synthase, EPOS D
FT	(AAV58577) "
FT	CDS
FT	54935..62254
FT	/tag- k
FT	/product= "Type I polyketide synthase, EPOS E
FT	(AAV58578) "
FT	CDS
FT	62369..63628
FT	/tag- l
FT	/product= "Cytochrome P450 oxygenase homologue, EPOS F
FT	(AAV58579) "
FT	CDS
FT	63779..64333
FT	/tag- m
FT	/product= "Orf 6 protein (AAV58585) "
FT	CDS
FT	complement (63853..64290)
FT	/tag- n
FT	/product= "Orf 7 protein (AAV58586) "
FT	CDS
FT	64363..64920
FT	/tag- o
FT	/product= "Orf 8 protein (AAV58587) "
FT	CDS
FT	complement (64287..64727)
FT	/tag- p
FT	/product= "Orf 9 protein (AAV58588) "
FT	CDS
FT	65063..65767
FT	/tag- q
FT	/product= "Orf 10 protein"
FT	CDS
FT	complement (65008..65874)
FT	/tag- r
FT	/product= "Orf 11 protein (AAV58590) "
FT	CDS
FT	complement (65871..66338)
FT	/tag- s
FT	/product= "Orf 12 protein (AAV58591) "
FT	CDS
FT	66667..67137
FT	/tag- t
FT	/product= "Orf 13 protein (AAV58592) "
FT	CDS
FT	67334..68251
FT	/tag- u
FT	/product= "Orf 14 protein (AAV58593) "
FT	CDS
FT	68346..68750
FT	/partial
FT	/tag- v
FT	/product= "Partial Orf 15 protein (AAV58594) "
FT	/note= "No termination codon given in the specification"
XX	
XX	
PN	W09966028-A2.
XX	
PD	23-DEC-1999.
XX	
PF	16-JUN-1999;
XX	
PR	18-JUN-1998;
PR	24-SEP-1998;
PR	05-FEB-1999;
XX	
XX	(NOVS) NOVARTIS AG.

PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX	
PI	Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX	WPI: 2000-097741/08.
DR	P-PsDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
XX	AAY58592, AAY58593, AAY58594.
PT	New isolated epoethione synthase genes, used for the recombinant
PT	production of epoethione for use in cancer therapy -
XX	Claim 14; Page 87-104; 174pp; English.
PS	
XX	This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC	EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC	the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
CC	and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC	formation. EPO F (AAY58579) is an epoethione macrolactone oxidase, and
CC	the proteins Orf 3 (AAY58582) and Orif4 (AAY58593) are thought to be
CC	involved in transport. Epoethiones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethiones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficiently exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethiones as anticancer agents, they are problematical to produce on a
CC	large scale. Epoethiones are too complex for industrial scale chemical
CC	synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC	poor yields of epoethiones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethiones in a heterologous host
CC	that is more amenable to fermentation.
XX	
XX	
SQ	Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
	Query Match 100.0%; Score 101; DB 21; Length 68750;
	Best Local Similarity 100.0%; Pred. No. 1.4e-21;
	Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1 CGCGGAGTGGCCGCTTACAGGGCGTGATCCATGCACCGAGCGCTCGATGATGCTGTC 60
DB	36162 CGCGGAGTGGCCGCTTACAGGGCGTGATCCATGCACCGAGCGCTCGATGATGCTGTC 36221
OY	61 TTGATGAGCAGACACCGAGCGCTTCTCGGGGTGTGGCA 101
DB	36222 TTGATGAGCAGACACCGAGCGCTTCTCGGGGTGTGGCA 36282
RESULT 2	
AAA29349	
XX	AAA29349 standard; DNA; 71989 BP.
AC	
XX	AAA29349;
XX	
DT	12-SEP-2000 (first entry)
XX	
DE	Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX	
XX	Epoethione; polyketide synthase; ePOA; ePOB; ePOC; ePOD; ePOE; ePOF;
KW	ePOK; ePOK; ePOK; ePOK; ePOK; ePOK; ePOK; ePOK; ePOK; ePOK; ePOK; ePOK;
KW	tubulin polymerization assay; anti-tumour; cytostatic; ds.
OS	
XX	Sorangium cellulosum.
XX	
FH	Key Location/Qualifiers

```
FT CDS 3..992 /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501 /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263 /*tag= c
FT /label= epaA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548 /*tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661 /*tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810 /*tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6155
FT /*tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493 /*tag= h
FT /label= epoB_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548 /*tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /*tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887 /*tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887 /*tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366 /*tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921 /*tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921 /*tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308 /*tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422 /*tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905 /*tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983 /*tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /*tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992 /*tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138 /*tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306 /*tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137 /*tag= x
FT /label= epoC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033 /*tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287 /*tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899 /*tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756 /*tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008 /*tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907 /*tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606 /*tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857 /*tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396 /*tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642 /*tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082 /*tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336 /*tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647 /*tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151 /*tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576 /*tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883 /*tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574 /*tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287 /*tag= ap
```

```
FT      /note= "encodes ER5"
FT      /tag= ag
FT      30321..30869
FT      /note= "encodes KR5"
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KS6"
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AR6"
FT      34107..34676
FT      /tag= au
FT      /note= "encodes DH6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes KR6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AR7"
FT      41341..41922
FT      /tag= bb
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match
Best Local Similarity 100.0%; Score 101; DB 21; Length 71989;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CGCGGAGTGGCGCTTACAGAGCGGTGATCCATGCACGCGAGCGCTGATGATGATGTC 60
      |||
DB      30550 CGCGGAGTGGCGCTTACAGAGCGGTGATCCATGCACGCGAGCGCTGATGATGATGTC 30609
QY      61 TTGATGACGACACGACCGCTTCTCGCGGGTCTGCA 101
      |||
DB      30610 TTGATGACGACACGACCGCTTCTCGCGGGTCTGCA 30650

RESULT 3
AA517367
ID      AA517367 standard; DNA; 33529 BP.
AC      AA517367;
XX
DT      12-MAR-2002 (first entry)
XX
DE      DNA sequence of S. cellulosum polyketide synthase cosmid, PKOS28-26.
XX
KW      Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
      acyl transferase domain; dehydratase domain; ketoreductase domain;
      acyl carrier protein domain; PKOS28-26; ds.
XX
OS      Sorangium cellulosum.
XX
PN      US6280999-B1.
XX
PD      28-AUG-2001.
```

```
PF      31-AUG-1998; 9805-0144085.
XX      22-JAN-1998; 9805-0010809.
XX      (KOSA-) KOSAN BIOSCIENCE.
XX      Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
PI      WPI; 2001-606536/69.
XX
PT      Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT      polyketide open reading frame encoding modules with one or more domains
PT      such as ketosynthase, acyl transferase and acyl carrier protein domains
PT
```

Claim 4: Fig 1; 72pp; English.

```
CC      The present invention relates to the isolation of novel Sorangium
CC      cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC      encoding them. The polyketide synthases include catalytic domains such
CC      as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC      ketoreductase domain and acyl carrier protein domain. A host cell
CC      comprising a PKS ORF (open reading frame) which encodes one or more
CC      more PKS domains is useful for producing polyketide synthases from which
CC      polyketides can be produced. The host cells are useful for constructing
CC      a library, where each individual colony of the library represents a
CC      colony with the ability to produce a particular PKS synthase and
CC      ultimately a particular polyketide. The polyketides produced by these
CC      colonies can be used collectively in a panel to represent a library or
CC      may be assessed individually for activity. Colonies in the library are
CC      also induced to produce the relevant synthases and thus to produce the
CC      relevant polyketides to obtain a library of candidate polyketides which
CC      can be screened for binding to desired targets such as receptors,
CC      signalling proteins, etc. The present sequence represents the DNA
CC      sequence of cosmid PKOS28-26 which encodes one or more domains of
CC      S. cellulosum PKS.
CC      Note: The present sequence is said to encode the functional domains
CC      of S. cellulosum PKS which correspond to domains or domain subsets of
CC      the amino acid sequences of ORP1 (AAU10700) and ORP2 (AAU10701).
CC      XX
```

Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

```
QY      3 CGCGAGTGGCGCTTACAGAGCGGTGATCCATGCACGCGAGCGCTGATGATGATGTC 62
      |||
DB      14155 GCGGCGCGCGCTTACAGAGCGGTGATCCATGCACGCGAGCGCTGATGATGATGTC 14214
QY      63 GATGACGACACGACCGCGCTTCTCGCGGGTCTGCA 100
      |||
DB      14215 GCGGCGCGCGCTTACAGAGCGGTGATCCATGCACGCGAGCGCTGATGATGATGTC 14252
```

```
RESULT 4
AAA92302
ID      AAA92302 standard; DNA; 31422 BP.
AC      AAA92302;
XX
DT      10-JAN-2001 (first entry)
XX
DE      S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW      Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
      multi-functional enzyme; polyketide; avermectin; veterinary drug;
      agrochemical; ds.
XX
OS      Streptomyces avermitilis.
XX
PN      Streptomyces avermitilis.
XX
PD      key Location/Qualifiers
      CDS 1..14646
```

```
FT FT /*tag= a
FT FT /note= "avermectin aglycon synthase protein"
FT CDS 14824..31422
FT FT /*tag= b
FT FT /note= "avermectin aglycon synthase protein"
XX XX WO200050605-A1.
XX PD 31-AUG-2000.
XX PE 23-FEB-2000; 2000WO-JP01041.
XX PR 24-FEB-1999; 99JP-0046961.
XX XX (KITA ) KITASATO INST.
XX PA Omura S, Ikeda H;
XX PI WPI; 2000-565458/52.
XX DR P-PSDB; AAB23751, AAB23752.
XX PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use
XX PS Claim 2; Page 134-203; 314pp; Japanese.
XX CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;
SQ Query Match 46.3%; Score 46.8; DB 21; Length 31422;
Best Local Similarity 67.3%; Pred. No. 7.1e-05;
Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 1 CGGCGAGTGGCCGTTACAGGCGTGATCATGCACGCCGAGCGCTCGATGATGCTGCG 60
Db 3863 CGGCGAGTGGCCGCTGGTGCTGTCATGCCGCGAGTGTGTGATGACGGGCTGT 3922
QY 61 TTGATGAGCAGACACGACGCGCTTCTCGCGGGTCTG 98
Db 3923 TGGAGGGTGTGTCGTCGACGCGTGTACAGGGGGTCTG 3960
RESULT 5
AAH79278
ID AAH79278 standard; DNA; 31422 BP.
XX AC AAH79278;
XX DT 04-DEC-2001 (first entry)
XX DE Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; veterinary drug; pesticide; ds.
XX OS Streptomyces avermitilis.
XX FH Key Location/Qualifiers
FH CDS 1..14646
FT FT /*tag= a
FT FT /product= "AAG65266"
```

```
FT FT 14824..31422
FT FT /*tag= b
FT FT /product= "AAG65267"
XX XX WO200162939-A1.
XX PD 30-AUG-2001.
XX PE 23-FEB-2001; 2001WO-JP01381.
XX PR 24-FEB-2000; 2000JP-0047405.
XX XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX PA (KITA ) KITASATO INST.
XX PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX DR WPI; 2001-582053/65.
XX DR P-PSDB; AAG65266, AAG65267.
XX PT New modified avermectin aglycone synthase derived from Streptomyces
PT PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides
XX PS Disclosure; Page 103-149; 257pp; Japanese.
XX CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoadacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoadacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermitilis genome.
XX SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
SQ Query Match 46.3%; Score 46.8; DB 22; Length 31422;
Best Local Similarity 67.3%; Pred. No. 7.1e-05;
Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 1 CGGCGAGTGGCCGTTACAGGCGTGATCATGCACGCCGAGCGCTCGATGATGCTGCG 60
Db 3863 CGGCGAGTGGCCGCTGGTGCTGTCATGCCGCGAGTGTGTGATGACGGGCTGT 3922
QY 61 TTGATGAGCAGACACGACGCGCTTCTCGCGGGTCTG 98
Db 3923 TGGAGGGTGTGTCGTCGACGCGTGTACAGGGGGTCTG 3960
RESULT 6
AAV05287
ID AAV05287 standard; DNA; 49377 BP.
XX AC AAV05287;
XX DT 21-MAY-1998 (first entry)
XX DE The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX KW Polyketide synthase; PKS; biosynthesis; soraphen; SOR; SOR; SOR;
KW SOR; biosynthetic module; beta-ketoadacyl synthase; acyltransferase;
KW ketoreductase; beta-ketone processing domain; cytosolic agent;
KW antimicrobial agent; phytopathogenic fungi; transgenic plant;
KW biological control; ss.
XX OS Sorangium cellulosum.
XX FH Key Location/Qualifiers
FH CDS 383..760
FT FT /*tag= a
FT FT /product= SOR
```

	/note=	"gene product highly homologous to the reductase domains of type I PKSs such as eryA from Saccharopolyspora erythraea"
FT		
ET		
CDS	/tag- b	927..19874
	/product= Sora	
	/note= "gene product is highly homologous to type I PKSs that are known to be involved in the synthesis of polyketide compounds"	
FT		
ET		
misc_feature	942..7115	
	/tag- c	
	/note= "module 1 of SORA"	
FT		
ET		
misc_feature	7203..12884	
	/tag- d	
	/note= "module 2 of SORa"	
FT		
ET		
misc_feature	13455..19616	
	/tag- e	
	/note= "module 3 of SORA"	
FT		
ET		
misc_feature	19871..46318	
	/tag- f	
	/tag- g	19871..46318
CDS	/product= SorB	
	/note= "gene product is highly homologous to type I PKS genes"	
FT		
ET		
misc_feature	19870..24556	
	/tag- h	
	/note= "module 1 of SorB"	
FT		
ET		
misc_feature	24638..30820	
	/tag- i	
	/note= "module 2 of SorB"	
FT		
ET		
misc_feature	30881..35446	
	/tag- j	
	/note= "module 3 of SorB"	
FT		
ET		
misc_feature	35528..40114	
	/tag- k	
	/note= "module 4 of SorB"	
FT		
ET		
misc_feature	40190..46318	
	/tag- l	
	/note= "module 5 of SorB"	
FT		
ET		
CDS	46851..47891	
	/tag- m	
	/product= SorM	
	/note= "gene product is homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin"	
FT		
ET		
US5716849-A.		
PN		
XX		
PD		
10-FEB-1998.		
XX		
PE		
14-DEC-1996;	96US-0764233.	
XX		
24-AUG-1993;	93WO-US07954.	
PR	08-JUN-1994;	94US-0258261.
PR	09-OCT-1996;	96US-0729214.
XX		
PA	(NOVS) NOVARTIS FINANCE CORP.	
XX		
PI		
Beck JU,	Hill DS,	Ligon JM,
Neff S,	Ryals JA,	Schupp T:
WIJ; 1998-	158369/14.	
XX		
DNA encoding Sorangium celulosum polypeptide(s) - used for, e.g., biosynthesis of soraphen useful as antimicrobial agent against phytopathogenic fungi		
PT		
X		
PS	Claim 2;	Columns 47-90; 64pp; English.
CC	The present sequence contains a cluster of genes that encode polyketide synthases (PKSs) that are involved in the synthesis of soraphens in Sorangium celulosum. The proteins encoded by the present sequence are	

CC	SorA, SorA, SorB and SorM. SorA and SorB contain biosynthetic modules
CC	which contain a beta-ketoacyl synthase, an acyltransferase, a beta-ketone
CC	ketoreductase and an acyl carrier protein domain, as well as a beta-ketone
CC	processing domains. S. cellulosum soraphens are useful as a cytosolic
CC	and antimicrobial agent active against phytopathogenic fungi.
CC	Soraphen-producing transgenic plants or biological control agents can
CC	also be produced, which may reduce crop losses and nutritional
CC	deprivation for local populations in many parts of the world.
SQ	Sequence 49377 BP: 7247 A; 19522 C; 14477 G; 8131 T; 0 other;
Query Match	43.6%; Score 44; DB 19; Length 49377;
Best Local Similarity	65.0%; Pred. No. 0.00056;
Matches	65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
Qy	1 CGCGGAGTGGCCGTTACAGGCGCATGCAGACCGGACGCTCGATGCTGTGC 60
Db	6277 CGTCGGAGTCCCGGTCAACGCGGCGTTTCAACAGCGAGAGACTCGCGACATGGCCTGA 6336
Qy	61 TTGATGACGACGACACCGACCGCTTCTCGCGGCTGCTGC 100
Db	6337 TCGGCGACATGACGCCGAGCGCATGCAGCGGCTCTTGC 6376
RESULT 7	
AAT80413	ID AAT80413 standard; DNA: 43280 BP.
XX	AC AAT80413;
XX	XX 27-FEB-1998 (first entry)
DE	Ty lactone synthase gene cluster.
XX	Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
KM	polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
XX	Streptomyces fradiae.
OS	
XX	Key Location/Qualifiers
FF	CDS 816..14243
FF	/*tag= a
FF	/transl_except= (pos: 816..818, aa: Met)
FF	/note= "ORF1 encodes protein shown in AAMW22601"
FF	14351..19945
FF	/*tag= b
FF	/transl_except= (pos: 14351..14353, aa: Met)
FF	/note= "ORF2 encodes protein shown in AAMW22602"
FF	20010..31199
FF	/*tag= c
FF	/transl_except= (pos: 20010..20012, aa: Met)
FF	/note= "ORF3 encodes protein shown in AAMW22603"
FF	31232..36067
FF	/*tag= d
FF	/note= "ORF4 encodes protein shown in AAMW22604"
FF	36249..41774
FF	/*tag= e
FF	/note= "ORF5 encodes protein shown in AAMW22605"
XX	EP791655-A2.
XX	
PD	27-AUG-1997.
XX	19-FEB-1997; 97EP-0301056.
PF	22-FEB-1996; 96US-0012078.
XX	
PA	(ELIL) LILLY & CO ELI.
XX	
PI	Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX	WPI: 1997-418046/39.
DR	P-PSDB; AAMW22601-W22605.

XX DNA encoding Streptomyces fradiae tyactone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
XX Claim 2; Pages 8-66; 220pp; English.
XX
CC This sequence represents the tyactone synthase gene cluster of the
CC invention. This sequence is also referred to as the tylg gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the synthesis of the polyketide tyactone, isolated
CC from Streptomyces fradiae. Tyactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
Query Match 40.8%; Score 41.2; DB 18; Length 43280;
Best Local Similarity 67.4%; Pred. No. 0.0041;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTGATGATGGTGTGC 60
DB 13093 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTGATGATGGTGTGC 13152
QY 61 TTGATGAGCAGACCCAGCCGCGCTTC 86
DB 13153 TCGACGGCGCTGACCCCGGAGCGATC 13178

RESULT 8
AADI7186
ID AADI7186 standard; DNA; 125401 BP.
XX
XX AADI7186;
AC 29-NOV-2001 (first entry)
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS 6337..34771
FT /*tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /*tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement (58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT note= "CDS does not include start codon"
FT complement (59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT note= "CDS does not include start codon"
FT complement (60238..61296)
FT /*tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"

XX
PN WO200159126-A2.
XX
XX 16-AUG-2001.
PD
XX
XX 08-FEB-2001; 2001WO-CB00509.
PF
XX
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIG.
PA (SMTF) SINTEF SPIFFEISEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIELEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVRIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI: 2001-557614/62.
DR P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
XX Claim 1; Page 188-254; 266pp; English.
PS
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 40.8%; Score 41.2; DB 22; Length 125401;
Best Local Similarity 64.9%; Pred. No. 0.0049;
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTGATGATGGTGTGC 60
DB 44786 CGCGGAGTGGCCGTTACAGGCGGTGATCATGACCGGAGCGCTGATGATGGTGTGC 44845
QY 61 TTGATGAGCAGACCCAGCCGCGCTTCGCGGGT 94
DB 44846 TCGACGGCGCTGACCCCGGAGCGGTTCGCCACGCT 44879

RESULT 9
AAE90038
ID AAE90038 standard; DNA; 5088 BP.
XX
XX AAE90038;
AC 06-AUG-2001 (first entry)
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic;
KW type I polyketide synthase; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..5088
FT /*tag= a

```
FT      /product- "type I polyketide synthase"
FT      /transl_except= "(1..3, aa: Met)"
XX
XX
XX      MO200140497-A2.
XX
XX      07-JUN-2001.
XX
XX      27-NOV-2000; 2000WO-FR03311.
XX
XX      29-NOV-1999; 99FR-0015032.
XX      07-JUN-2000; 2000US-0209800.
XX
XX      (AVERT ) AVENTIS PHARMA SA.
XX
XX      Jeanin P, Pernodet J, Guerlineau M, Simonet P, Courtois S;
XX      Cappellano C, Franco F, Raynal A, Ball M, Sezonov G, Tuphille K;
XX      Frostegard A;
XX
XX      WPI; 2001-374849/39.
XX      P-PSDB; AAB83975.
XX
XX      Collection of nucleic acids from environmental samples, useful for
XX      identifying e.g. genes encoding polyketide synthases and derived
XX      antibiotics
XX
XX      Claim 35; Page 309-311; 356pp; French.
XX
XX      The specification describes a method for the preparation of a collection
XX      of nucleic acids from organisms in a soil sample. The method comprises
XX      milling a dried sample to produce microparticles; suspending these in
XX      liquid buffer; extraction of nucleic acids from the microparticle;
XX      passing nucleic acid-containing solution through a molecular sieve;
XX      passing nucleic acid-enriched fractions through an anion exchange
XX      chromatography material; and recovering fractions containing purified
XX      nucleic acids. The nucleic acids are sources for sequences that encode
XX      either operators involved in a metabolic pathway (specifically polyketide
XX      synthetase) or polypeptides, particularly for production of therapeutic
XX      or agricultural compounds, especially polyketide antibiotics. AAF90034-39
XX      represent open reading frames (ORFs) of the coding strand of cosmid
XX      a26g1, and encode type I polyketide synthases.
XX
XX      Sequence 5088 BP; 838 A; 1622 C; 1679 G; 949 T; 0 other:
XX
XX      Query Match      40.0%; Score 40.4; DB 22; Length 5088;
XX      Best Local Similarity 65.6%; Pred. No. 0.0049;
XX      Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
XX
XX      11 GCGGTACAGGCGGTATCATCCAGCGGCTGATGATGTGCTTATGACA 70
XX      ||||| | | | | | | | | | | | | | | | | | | | | | | |
XX      DB 4239 GCGGCTGCGTGGAGTGTGATGCCAGCGGCTGATGAGGAGTATGACAGACA 4298
XX
XX      71 GACCAACGACCGGCTCTCGCGGGGTGCTGGC 100
XX      ||| | | | | | | | | | | | | | | | | | | | | | |
XX      DB 4299 GACGTGGCGCTCTTTCGAGAGAGGTCTGGC 4328
XX
XX      RESULT 10
XX      AAF24892
XX      ID AAF24892 standard; DNA; 20394 BP.
XX
XX      AAF24892;
XX
XX      20-APR-2001 (first entry)
XX
XX      Pimaricin biosynthesis associated polyketide synthase gene.
XX
XX      Polyketide synthase; oxidative modification; metabolite; antibiotic;
XX      anticancer; pimaricin; ss.
XX
XX      Streptomyces natalensis.
XX
XX      Key Location/Qualifiers
XX      CDS 1..20394
```

```
FT      /*tag= a
FT      /product= "polyketide synthase"
XX
XX
XX      MO200077222-A1.
XX
XX      21-DEC-2000.
XX
XX      14-JUN-2000; 2000WO-EP06227.
XX
XX      14-JUN-1999; 99EP-0201893.
XX
XX      (STAM ) DSM NV.
XX
XX      Martin JF, Aparicio JF, Colina AJ;
XX
XX      WPI; 2001-080693/09.
XX      P-PSDB; AAB31558.
XX
XX      New polynucleotides encoding enzymes involved in the biosynthesis of
XX      pimaricin, useful for modifying the biosynthesis of pimaricin and in
XX      the synthesis of new compounds
XX
XX      Disclosure; Page 53-80; 116pp; English.
XX
XX      The present sequence encodes a polyketide synthase which is associated
XX      with the biosynthesis of pimaricin. The polyketide synthase polypeptide
XX      is useful for the oxidative modification of a methyl group of a suitable
XX      compound, e.g. a bioactive compound including a secondary metabolite,
XX      antibiotics and anticancer agents. Recombinant cells comprising the
XX      gene are useful for the production of pimaricin. The polyketide synthase
XX      polynucleotide may be over expressed in Streptomyces, leading to an
XX      increase in the biosynthesis of pimaricin, as a source of primers for
XX      amplification reaction and as probes.
XX
XX      Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other:
XX
XX      Query Match      40.0%; Score 40.4; DB 22; Length 20394;
XX      Best Local Similarity 63.3%; Pred. No. 0.0063;
XX      Matches 62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
XX
XX      1 CGCGGAGATGCGCGGTACAGGCGGTATCATCCAGCGGCTGATGATGTGCTG 60
XX      ||||| | | | | | | | | | | | | | | | | | | | | | | |
XX      DB 3797 CGCGGACCTGCGCGGTACCGGCGGTCTCCACGCCGCGGTACGAGAGCGGCTAC 3856
XX
XX      61 TTGATGACAGACACCGACCGGCTCTCGCGGGGTGCTG 98
XX      ||||| | | | | | | | | | | | | | | | | | | | | | |
XX      DB 3857 TGGACCGGCTACCCCGGACCGCTTCGAGACCGTACTG 3894
XX
XX      RESULT 11
XX      AAF90033
XX      ID AAF90033 standard; DNA; 34071 BP.
XX
XX      AAF90033;
XX
XX      06-AUG-2001 (first entry)
XX
XX      Nucleotide sequence of cosmid a26g1 (coding strand).
XX
XX      Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX      Synthetic.
XX
XX      MO200140497-A2.
XX
XX      07-JUN-2001.
XX
XX      27-NOV-2000; 2000WO-FR03311.
XX
XX      29-NOV-1999; 99FR-0015032.
XX      07-JUN-2000; 2000US-0209800.
XX
XX      (AVERT ) AVENTIS PHARMA SA.
```

XX Jeanin P, Perodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
XX Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticles;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand, which encodes different
CC polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
Query Match 40.0%; Score 40.4; DB 22; Length 34071;
Best Local Similarity 65.6%; Pred. No. 0.0069;
Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 11 GCCGTTACAGGCGGTCGATGCATGCAGCCGAGCGCTCGATGATGCTTGATGACA 70
DB 28920 GCCCGTGGTGAGTCGTCGATGCAGCGCGCTGCTCGATGCAGGCGTATGACAACA 28979
OY 71 GACCACCGACCGCTTCGCGGGTCTGGC 100
DB 28980 GACGTGGGCTGCTTTCGAGAGAGTCTGGC 29009

RESULT 12
AAF90032/c
ID AAF90032 standard; DNA; 42717 BP.
XX
AC AAF90032;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS Synthetic.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR0311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Perodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX WPI: 2001-374849/39.
XX

PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
XX Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticles;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC encodes different polyketide synthases.
XX
SQ Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
Query Match 40.0%; Score 40.4; DB 22; Length 42717;
Best Local Similarity 65.6%; Pred. No. 0.0072;
Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
OY 11 GCCGTTACAGGCGGTCGATGCATGCAGCCGAGCGCTCGATGATGCTTGATGACA 70
DB 5200 GCCCGTGGTGAGTCGTCGATGCAGCGCGCTGCTCGATGCAGGCGTATGACAACA 5141
OY 71 GACCACCGACCGCTTCGCGGGTCTGGC 100
DB 5140 GACGTGGGCTGCTTTCGAGAGAGTCTGGC 5111

RESULT 13
AAH52062
ID AAH52062 standard; DNA; 4851 BP.
XX
AC AAH52062;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 116.
XX
KW Drug target; growth; organism viability; characterisation; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX WPI: 2001-329193/34.
DR P-PSDB; AAG81211.
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyses a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences
XX
PS Disclosure; Page 131-133; 207pp; English.
XX

CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.

SQ Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

Query Match 39.0%; Score 39.4; DB 22; Length 4851;
Best Local Similarity 62.9%; Pred. No. 0.01;
Matches 61; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 GGGGAGTGGCGCTTACAGGGCGTATCGATGAGCGGAGCGCTCGATGTGTGCT 61
DB 3915 GGAGTACCCCGCGGTGGGGGGGTATTCATGCCCGCGCTGCTCATACGACGTGAT 3974
QY 62 TGATGAGCAGACGACCGACCGCTTCTCGCGGTGCTG 98
DB 3975 CACCTCGTTGACACGCGACCGCATCGATGAGGTGTTG 4011

RESULT 14
AA168063
ID AA168063 standard; DNA; 552 BP.

XX AA168063;
XX
XX 13-MAR-2002 (first entry)

DE Epothilone PKS KR domain nucleotide sequence.

XX
XX
XX Cystobacterinease; recombinant; polyketide synthase; PKS; polyketide;
KW epothilone; p11A gene; cytosolic; antiparasitic; antiarthritic;
KW antitartaric; antitartaric; antitartaric; neuroprotective; vasotropic; ds.
XX
XX Synthetic.

OS
FH Key Location/Qualifiers
FT CDS 1..552
FT /*tag= a
FT /note= "KR domain of extender module 6"

XX
XX
XX WO200183800-A2.
XX
XX
XX 08-NOV-2001.
XX
XX
XX 26-APR-2001; 2001WO-US13793.
XX
XX
XX 28-APR-2000; 2000US-0560367.
PR 14-SEP-2000; 2000US-232696P.
PR 21-DEC-2000; 2000US-257517P.
PR 03-APR-2001; 2001US-0825856.
PR 03-APR-2001; 2001US-0825876.
PR 13-APR-2001; 2001US-269020P.
XX
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.

XX
XX
XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosla C;
PI Lau J, Liscardi PJ, Regentin R, Santl D, Tang L;
XX
XX WPI; 2002-075167/10.
DR P-PsDB; ABB07167.
XX
XX Recombinant host cells useful for producing polyketides e.g. epothilone
PT or its derivatives, comprises a recombinant expression vector encoding

PT a heterologous polyketide synthase gene -
XX
XX Example 11; Page 164-165; 221pp; English.

PS
XX The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of the
CC KR domain of extender module 6 of the epothilone PKS. Inactivation of
CC this domain results in a novel PKS capable of producing a 9-keto-
CC epothilone analogue.

SQ Sequence 552 BP; 70 A; 176 C; 214 G; 92 T; 0 other;

Query Match 38.8%; Score 39.2; DB 24; Length 552;
Best Local Similarity 62.0%; Pred. No. 0.0079;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCGCTTACAGGGCGTATCGATGAGCGGAGCGCTCGATGTGTGCTG 60
DB 233 CGTGGGGGATGCCGCTCCGCGGTGCTATCATCGCGCGGTATTCGAGCAGCGGCTGCG 292
QY 61 TTGATGAGCAGACGACCGACCGCTTCTCGCGGTGCTGCG 100
DB 293 TGATGAGCAGACCGACCGCGGTGCTCGCGGTGCTGCG 332

RESULT 15
AA168064
ID AA168064 standard; DNA; 552 BP.

XX AA168064;
XX
XX 13-MAR-2002 (first entry)

DE 9-keto-epothilone PKS inactive KR domain nucleotide sequence.

XX
XX
XX Cystobacterinease; recombinant; polyketide synthase; PKS; polyketide;
KW epothilone; p11A gene; cytosolic; antiparasitic; antiarthritic;
KW antitartaric; antitartaric; antitartaric; neuroprotective; vasotropic; ds.
XX
XX Synthetic.

OS
FH Key Location/Qualifiers
FT CDS 1..552
FT /*tag= a
FT /note= "mutated KR domain of extender module 6"

XX
XX
XX WO200183800-A2.
XX
XX
XX 08-NOV-2001.
XX
XX
XX 26-APR-2001; 2001WO-US13793.
XX
XX
XX 28-APR-2000; 2000US-0560367.
PR 14-SEP-2000; 2000US-232696P.
PR 21-DEC-2000; 2000US-257517P.
PR 03-APR-2001; 2001US-0825856.
PR 03-APR-2001; 2001US-0825876.
PR 13-APR-2001; 2001US-269020P.
XX
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Arslanian RL, Ashley G, Frykman S, Julien B, Katz L, Khosia C;
PI Lau J, Licardi PJ, Regentin R, Santi D, Tang L;
XX WPI; 2002-075167/10.
DR P-PSDB; ABB07168.

XX Recombinant host cells useful for producing polyketides e.g. epothonone
PT or its derivatives, comprises a recombinant expression vector encoding
PT a heterologous polyketide synthase gene

PS Example 11; Page 165-166; 221pp; English.

CC The invention provides a recombinant host cell, of the suborder
CC Cystobacterineae comprising a recombinant expression vector encoding a
CC heterologous polyketide synthase (PKS) gene and produces a polyketide
CC synthesized by the PKS enzyme encoded on the vector. An epothonone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a diketide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothonones, including new epothonone derivatives, but also
CC other polyketides. Methods of purifying the epothonone derivatives are
CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or restenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of a
CC mutated and inactive KR domain of extender module 6 of the novel 9-keto-
CC epothonone PKS of the present invention.

XX Sequence 552 BP; 70 A; 177 C; 212 G; 93 T; 0 other;

Query Match 38.8%; Score 39.2; DB 24; Length 552;
Best Local Similarity 62.0%; Pred. No. 0.0079;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY 1 CGGGGAGTGGCCGTTACAGGGGCGGATGCATGCAGCCGCGGCGTGCATGATGATGTC 60
DB 233 CGTGGGGATGCCCTCCGCGGCTGCTTCATGCGGCGGATCCCTGACGAGGGGCTGC 292
OY 61 TTGATGAGCAGACCCAGCCGCTTCTCGCGGCTGCTGC 100
DB 293 TGATGACAGCAACCCCGCGGCTTCGCGGCGGATGATGTC 332

Search completed: November 6, 2002, 12:14:42
Job time : 260 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:51:49 ; Search time 443.625 Seconds

(without alignments)
6625.829 Million cell updates/sec

Title: US-09-724-876-2_COPY_30550_30650

Perfect score: 101

Sequence: 1 cgcgcgagtcgcccgtaccag.....gtcttcgcgcggtcgtgca :01

Scoring table: IDENTITY_NNC

Gapop 10.0 , Gapect 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenBml: *
1:  gb_da: *
2:  gb_htg: *
3:  gb_in: *
4:  gb_om: *
5:  gb_ov: *
6:  gb_pat: *
7:  gb_ph: *
8:  gb_pl: *
9:  gb_pr: *
10: gb_ro: *
11:  gb_sts: *
12:  gb_sy: *
13:  gb_un: *
14:  gb_vi: *
15:  em_ba: *
16:  em_fun: *
17:  em_hum: *
18:  em_in: *
19:  em_mu: *
20:  em_om: *
21:  em_ov: *
22:  em_ov: *
23:  em_ph: *
24:  em_pat: *
25:  em_pl: *
26:  em_ro: *
27:  em_sts: *
28:  em_un: *
29:  em_un: *
30:  em_htg_hum: *
31:  em_htg_inv: *
32:  em_htg_other: *
33:  em_htg_mus: *
34:  em_htg_pln: *
35:  em_htg_rtd: *
36:  em_htg_man: *
37:  em_htg_vrt: *
38:  em_sy: *
39:  em_htgo_hum: *
40:  em_htgo_mus: *
41:  em_htgo_other: *
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	68750	1 AF210843	AF210843 Sorangium
3	101	100.0	68750	6 ARI93029	ARI93029 Sequence
4	101	100.0	68750	6 ARI93029	ARI93029 Sequence
5	101	100.0	68750	6 ARI99551	ARI99551 Sequence
6	101	100.0	68750	6 ARI99559	ARI99559 Sequence
7	101	100.0	68750	6 ARI99567	ARI99567 Sequence
8	101	100.0	68750	6 AR201097	AR201097 Sequence
9	101	100.0	68750	6 AR208671	AR208671 Sequence
10	51.6	51.1	33529	6 ARI72654	ARI72654 Sequence
11	47.8	47.3	49736	1 AF3166425	AF3166425 Sequence
12	46.8	46.3	31422	6 AF319998	AF319998 Stigmatel
13	46.8	46.3	64957	1 E38021	E38021 Avermectin
14	46.8	46.3	64957	1 AB033367	AB033367 Streptomycin
15	44	43.6	49377	6 I88042	I88042 Sequence
16	43.6	43.2	67523	1 SCU24241	SCU24241 Sorangium c
17	43.6	43.2	66808	1 AF016585	AF016585 Streptomycin
18	42.6	42.2	30000	6 SAU421825	SAU421825 Stigmatel
19	41.8	41.4	27522	1 AX250262	AX250262 Sequence
20	41.2	40.8	43280	1 AB070942	AB070942 Streptomycin
21	41.2	40.8	125401	1 SFU78289	SFU78289 Streptomycin
22	41.2	40.8	125401	1 AF263912	AF263912 Streptomycin
23	40.4	40.0	5088	6 AX211739	AX211739 Sequence
24	40.4	40.0	20394	1 AX153795	AX153795 Sequence
25	40.4	40.0	20394	1 SNA132222	SNA132222 Streptomycin
26	40.4	40.0	24225	6 AX067996	AX067996 Sequence
27	40.4	40.0	34071	6 SC2C4	SC2C4 Streptomycin
28	40.4	40.0	34071	6 AX153790	AX153790 Sequence
29	40.4	40.0	39314	1 SGR300302	SGR300302 Streptomycin
30	40.4	40.0	42717	6 AX153789	AX153789 Sequence
31	40.4	40.0	84985	1 SNA278573	SNA278573 Streptomycin
32	39.4	39.0	113193	1 AF357202	AF357202 Streptomycin
33	39.4	39.0	16096	1 AE007124	AE007124 Mycobacter
34	39.2	38.8	552	1 MTCY2461	MTCY2461 Mycobacter
35	39.2	38.8	552	6 AX403009	AX403009 Sequence
36	38.8	38.4	16767	6 AX089464	AX089464 Sequence
37	38.8	38.4	28736	6 AX089421	AX089421 Sequence
38	38.8	38.4	50000	6 AX089417	AX089417 Sequence
39	38.8	38.4	53789	1 AMM223012	AMM223012 Amycolato
40	38.8	38.4	53789	6 A69720	A69720 Sequence
41	38.8	38.4	80161	1 AV007564	AV007564 Saccharop
42	38.8	38.4	80161	6 ARI65018	ARI65018 Sequence
43	38.8	38.4	90445	1 AF040570	AF040570 Amycolato
44	38.2	37.8	50000	6 AX089420	AX089420 Sequence
45	37.2	36.8	2700	6 AX211733	AX211733 Sequence

ALIGNMENTS

```
RESULT 1
LOCUS AF217189
DEFINITION AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189
KEYWORDS GI:7453554
SOURCE
ORGANISM Polyangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)
```

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epochlone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epochlone biosynthetic gene
cluster from *Sorangium cellulosum*
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kusan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1. 58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
<1. .992
/note="orf1"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VNGHEAPADFGVPRVLLYDNLKNAVVERHGDATIRFHTLLAL
SADYREPRVAVARGNEKRGVERALRYRGEFFEARAYADLGRQATWNTSSNAL
DSRWEDRARTRQAQFDDRSYLLRHDPDPDHERVEVEGKTPYARPDLDYVPH
DSTRRLVYADLSQYRIADGNOIVATHYRSMRDGGOITOEPIHOLRIYDEKRRARH
GIDRLARRASQATRLRTAERGDNVGAIALRLQILDVGAALDEALVEYLERNTI
HGAVRQVYDRRRSERHLPPIVPTKGEHALVYTPHSLTTIDLKADPT"
989. .1501
/note="orf2"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTETKDKRLKSLGFLGACMEQLADKRWLREVLIAIEERE
RHRSLSERLKNRSYAAAFKPMTPDPSMPKIDREAVDOLYDSRYADLLFEVYTRAYD
AOKPLLSLNKAFADMGVYPAACVYTLVDRLVHRAVEITEASRYLKEAKELNAT
RTKORRTKH"
1998. .6263
/gene="epoA"
1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epochlone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADPRERAAEDPTAIYAGAGRLPGCVTIDSGFWLLRGSRTV
GOVPAERMDAAMFPDLDAPKCTPTFRASPLSDVACFPASFFGTSPPRALMDGPRH
LLEVCWEALENAAIAIPSAIVGTETGVFTGIPSEVEALPRATYSAEIDAGLGTMT
PSVGAGRISYGLRGPCVAVDTAVSSSLVAVHLAQSLRSGECSTALMGRLVLMISP
STVWLSTKTRLATDGRCAKFAEADGFGRGCAVAVVYLRSGAARDRLIAVLRG
SAIINDGASGGLTVNGSSEIIVLRALADGCAASVGVYEHGHTLGDPIEIOA
LNAVGLGDVATPLLIGSVKTNLIGHPVEYASGITGLIKYVLSLHQQTIAHLHAQALN
PRISMGDLRTYTRARTPPDPNTPRRAGVSFGMSGTAHVYLEAPATCTPAPE
RPAELLVLSARTAAALDAHAARLDHLETPSQCLGDAVASFATTSSAEMHRLAAVAT
SSEGLRAALDAAGQCTPGVVGRIADSSRGKIAFLFTGOGAQTLMGRLYDVMPAF
KEAPDLCVRLTNOELDRPLREYMAEPASVDALADQTAFTQPALTFEYALAAALRS
WGPEELVAGHSIGELVACVAGVSELEDAVLVAARGRLMALPAGGAMVSIAPBA
DYAAAIVAPHAASVSTAAVNGPDQVYIAGAGOPVHAIAAAMARGARTKLAHSHAFS
PLAMPMLEHFGVAVSASTYRPSIYVLSNLSKAGIDEVSSFGYVYRIHAREVYRRADG
VVALHAAGAGTFVEGPKRSTLLGLVPACLPDAPRALASSRAGREDEPVALGLGLW

AVGLVSNAGLFPSSGGRVPLPTYPWQRRERYIDTKADDAAGDRRAPGAGHDEVEKG
GAVRGDRRSARLDHPPEPSGREKVEAEADGPRFLIEDPEGLVRLVRYTERAPGL
LGEVEIYVPAAGLSPENDVOLALGMPDDLPAGPNPFLILGGCACRIVAGGVNGVLV
VGOPVTAISAGAPATHVTSAALVLPAPALATATRAAMPVAYLTAAWYADSLARLP
GERVLHAATGSGVGLAAYVMAQHVGAENAVTIGTPEKRLYLESLYRYVDSRSRFP
ADVRAWTGGSEVDVYNLSGELIDKSEHLSHGFVEYLGKRDYADQDLGRPLR
NLSFSLVDKGMLEPRARVALFELLGLIAAGVTPPPVIAFLIARVADFRMAQ
AOHIGKIVLTLDPEVOIRIPTHAGGPGSTGDRDLDRLASAAPAARAALAEPLRTQ
VOVLRTEPEIKGEALFRLMDSILMAVELNRIEASIKLSTFTSTSNIALLT
ONLIDLATATLSIENVAENLRAQVSDPVSSGAGQDMETIAL"
6260. .10492
/gene="epoB"
6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epochlone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLNLEHOGVTKLAADDERLOIQAQKNAIPLNLLARISEH
KSTILMLKQRLPAESIYAPARERHYPPFLTDIOGSYMLGRGATTVSGTHAYREYD
CTDLDAVRLSRARVVARHDMLRHATTPDMQVLEPKVADIEITDLRGDRSTREA
RLVSLDAMASHRIYDTERPPLVAVARLDEQOTRLVSLIDLINDLSLIIFFKDWL
SEYEDPETSIPVLELSYRDYVLALESRRKSEHQSMDYKRRVAVELPPPMIPKAD
PSLREIRFRHTEOMLPDSMSRLKQVREGLTPGVILAFSEVIGKMSASPRFTL
NITLERNRLHPHRVDITGDFTSMVLLDITDRDSFEQRAKRIQOLEMEANDDVS
GIENQREARVIGIQRGALFPVYIYLSALNOQVYGTSLQRCSTPYTISQTPQLLDH
QLEHGDVLYLMDYVDGVFPDLDDMLAEIYATLRLRLEPSEKQMSCLPPQLE
AASANENSLSEHTLHGLFAARVQLEPMQAAVARSATTLVEELRSRSRLGARLR
BOGARNTLVAAVMEKMGEOVAVAVLAEVLAAYVIDADLPAEBLHVLLHGEVLY
LQPMIDGKLSMPGQIQRLLVDAEGEGDGLPMPLOTPDGLVYITSGSTELPK
GVMIHGRGAVNTLDOIINERFEIGPDRVYLSLSEDSYVYFGLIAGGTIVPDA
SKLPDPAHAAALIEBEKYVYVMSVPAIMSLVYEHGDRDSLARSLTSLSGMDIVP
GLPGELOARIPGVYSISLGGATEASTWSIGTYRVAVNDPMASITPIGRPLRDTRFVLD
EALPEPRVWPGQLYTGGVLAGLYWRDEKRNSEFLVHPETGERLYETGDDGRFLPD
GNIERFREDNOIKLRGYNEVLEGEIETLKSHPNVDATYVAGDANKLLALAVVP
EGTRRAAEQDASLATERVDARAHAAKADGLSDGRVOFKLARHGLRDLDDKPAVVD
TGLVPREAGLDVYARRSRVYRLEAPIPEVFERGLSCISVEPGGALPKRRYPSAG
SPYVQTYAVYAKSGRIEGYDEGEFYHYHSHGIRGHAHVONPVPVEAA
FGLVYGRIDAIESLYGSLSRFCLLEAGYMAQLMEQAPSCNITVCYGGQDFEQVR
PVLDLRHSDVYVHGMIGGRVDFQVCTLGDDSPRRATYTGAPGQHQAPADLEIR
FLRTKRIPEYMPVTEVEYDLPLTNGKVRALREKRTDTSPPRSHGTAAPDALEEI
LVAAYVREYGLLEVGLQOSFVLDGATSIHIVYRSLQKRLDREIATITELFQYPMIGS
LASGLRDKSKDLQGNPMDQVREARRKGRRS"
10639. .16137
/gene="epoC"
10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epochlone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEDSSAIAIVIGMSGRFPGARDDDEFERNLRDGEAVQRESQ
ELAAGCPALVLDPNYVAGSVLVDVRFDAAPFGISPREAELMDPHRIEMCEWE
ALENAGQPTATIEGSIYVYAGANMSSTLYLNRHEPRAARWPMQOTILGNDKIYAT
HVSYRLNLEGPSISVQTAQSTSLVAVHLAKSLDRDECMALAGSITVLRPIRATYV
AGGIGFSPDGHRAVDANKANGIMONGVYLLKPLDRALSGDGVRAVYILGASANN
GARKIGETPSEVGOAOLAMEALALAGYEARSIOYIEHGTGTLGDALETPALARVE
DDBASTRRSCAIGSVYTGHLSEAGIAGLTKYVLALEHROLPSLNEESPNSIDE
ASSPYYVNTSKDMQNTGPPRRAGVSFSGITGTAHVVYLEAPAKLPAAAPARSAEL
EYVSAKSAALDAAARLDHIOAHQGLSLGVATSLATTSRPMHRLAAMAPSEAL
RGCLDAAGQPPPAVNRGCSPPGNVAVFVFPQGSOWMGWGLAAEPEFIAL
SACDRAIOEAGMSLAEIADAGSSQIETRDVDPVYFALVAPAAALMRSGVGPDV
YGEHMGVEYAAHVAGALSIEDAVAILICRSRLRRIISGOGEMAVTELSLEAKARL
VIEDSVYVNSPSTYLSGEPALIGEYLSLANKGYCRRVAKYDVASHSPOVPLR
BDLLAAGLTPRAAIVPMRSVYVGAAPGLGAYNMNNLRQVRAEAYVOALQIG
GHGLVEASPHRILTTSVEEMRRAQKAGAVGSIYRCODEPRALVELGALMAQGT
YVWGRLPAGGRVPLPTYPMQRRERYWIEAPKASAGRGVRAAGRHPLDGLTST

QSTRMETTLDLRLPWLGDHROGAVFPGAATLEMAISSGAEALDGPLOITDVA
LAELAFAGDAVALVOYTTTEOPSGRLOFOIARSPARGHSPFRHANGALBVERTE
VPAGTISAVARARLOASMPAAATAYELTETMLOQCPAFOGAIETELMREGSEALRRLP
DAAGSAEIRLHPALDLACFOVOSLFRGGEATPWEVEGSEKLRORPGEIMCHA
RVNHRQTPDROGADFWYDSSGAVVAORLPJGVREEDDWLELEMER
AAVGTAKVNAGRMILLGGGGIGALRSMELAGAVVAHAEESNTSAAGVALLAKAF
DGOAPTAVHLGSLDGGELDPGLGACALDAPRSADVSPALDPLALRGCDVLMTV
OALMGFRDAPRLMLTRGNAVAGAVSTOAPLGLGIVTIMEHDLRCARDLID
PTPRDGLGALLAEALDADAEEVALRGECVARIYAROPETPRRGITESCVPITDVI
IRADSTVLVTGGLGSLVAGMLAERAGHLVLVGRGSAVSDEORAAVLEAGAR
VTAKADVADRAOERILREVTTSGMPLRGVHAAGLIDDLMOQTPARARKVAPK
VOGALHHLATFREAPLSEFVLVAGVGLGSPGGONYAAATFLDLAHNRAGOLIPA
LSVDMGLEAEVMAAOEDRGARLVYSGRMSLTDEGSLATALLLESRGVOYGVAPV
PLMWELYPAAASSRMLSRITLTAHRSAGPAGCDLIRLAAAPESRSGLEPILR
AOISQVRLPEKIEVDARPLTSLGNSLMGELRRRIEMAGITVPLMLTTPVVA
LSGHLARACCAAPVESPHTTADSAVEIEEMSODDLQLIAKFRALIT"

gene
16134. .37907
/gene="epod"
/product="epod"
/note="type I polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTKPTAQONPLKQAIIIQREERLAGLAQALELTERPIAV
GIGRFGGADAPAFWELIDAEKDAVOPDLRMAVGVAVEVPHVAGLITPIDC
FDAAFGISPREASLDPOHRLLEVAEGLEDAGIPRSIDGSRTPGAGVGLADYA
RTVARPREEDASATGNMLSLAAGSLTYLGLQGPCITVDTCSSLVIAHLACS
LRAGESDLALAGVSTLSPDMMEAAARTALSPDRCRTFDASANGFVREGGCIAY
LKRLSDAQRDRIHALIRGSAIINHDSGTCTIAPNTVAQETVLRALRSAAHEGAY
DYVEHGTGSLDPTVEALRATVGPARGSGTRCVLGAVTNIGHEAAGVAGLIK
AALSILHERIPRNLFRTLNPRIRLEGSALALATEPVPRTDRFRFVGSFSGST
NAHVLEAPAVELEMPAPERSAELIYVSGSEGALDOAARLREHLMHEGLIGDYA
AFSLATRESAMTHRLAVAVTSREGILLASVAGOOTPAGAACRCIASSRCKLILFT
GOGAOTPGMGICAMPAREAFRCVTLFDRELDRPLREVMANEGASATLIDOT
AFQOPALFAVEYATLALMRNSGVEPRLVGSIGELVACVAGVSELDGVRILVADP
RLMGSLGAGMVASIGAPEAEVAAVAPHAAMVSTIAANGPEDQVIAVEAOVQIAA
GFAARGVTRKHLVSHAFSPHLEPMELEEFGRVAAVSVYRSPVSLVSNLSGKVVYDDE

Query Match
Best Local Similarity 100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGAGTGGCCGTTTACAGGGCGTATCCATGACAGCCGAGCGCTCGATGATGCTGTC 60
|||||
Db 30550 CGCGGAGTGGCCGTTTACAGGGCGTATCCATGACAGCCGAGCGCTCGATGATGCTGTC 30609
|||||
Qy 61 TTGATGACGACGACGACCGCCCTTCTGCGGGGCTGTCGCA 101
|||||
Db 30610 TTGATGACGACGACGACCGCCCTTCTGCGGGGCTGTCGCA 30650
|||||

RESULT 2
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangiineae; Polyangiaceae; Polyangium.
REFERENCE
AUTHORS
Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milmanow, M.,
Nowak-Thompson, B., Engel, N., Toupet, C., Strattmann, A., Cyr, D. D.,
Goriach, J., Mayo, J. M., Hu, A., Goff, S., Schmid, J., and Ligon, J. M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90

JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662635
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar, I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd., P.O. Box
12257, Research Triangle Park, NC 27709, USA
location/Qualifiers
1. .68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1. .1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF62904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSARVLDGHRATERHYLAERGIEDLRALREH
LRIDGSPSFHMCIGDLTVLELADHDPLASISFRHARSLEHPMTSDAMLVDPALV
RWLAARGAPLREYEERERARTRAOEARLMLAAAPCPAPDLPREDANDGLPG
MSPEVAERLRASATPELACALLAMLTGAGPMGSGPAYEMLPENLLGGLPT
AIAASARCTSEALRGAARLPASWEVSSKSGOENPEALMRRLPTIVRAGNADN
LSREERATIAAEVRLARQAPAPFAGAGLVAGVSSSGRLSGLVTTGDALISGDND
LVMPQGRISVPLVLAGDPPEELAPPLSOMLFAHANAAGTISKVLEGSPLVMARN
OARPSMLVHANGFAMVNOAMVPBERGAPVVOYSTIMEHEPTPCRLHBPAGSAS
LACDEEHYKWEISLGRLEMRHPHRRGASRPAAYLGEHP IAAWTWPSLTINATHVL
WADPPRRALIVDQKRGVEPVLVETRRHPAVVSESDIDITALGQDPDSRDMVHEIR
SGASIVAVDQROLMDRPDMVLTNRGLFEFTINDRLITARS"
complement(1900. .3171)
/note="ORF2; similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to DD-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF62905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPCALGAEDEEGPDGAGALAPAREVMAEVA
AGOMGAVMLVARGDDVDAVGVTELGSAPMRDTFRASMTKATVATVMMLE
EGKIDLSDPVRMLPDELARRVARIQSDIDETVPAERPTVRDMLPMTMGISFPA
SSPIORATIDELGVNAQPVPTPHCPDMEIRRLGLPLMHQPGAMWNTSIVQVYL
VGRADQGFDAFVRRILAPLGMRTDDEVADKLARVAGCGFTDEQGTGKTRMDR
GAESAIVASPPAPFPGAGLIVSTVDYLLFABMLNGVHGBRRLISAAYEMRADHL
TPQAKAASFPGFEFTGMYGMAVVPADAVSEVPGRYWMDGFGTSWINDGREL
IGIVTOSAGFLFSGALERFMRSVVATESA"
3415. 5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF62906.1"
/db_xref="GI:6724240"
/translation="MHGITROYVLSLTALILVTARASELIARLQPEVLELFC
GVIVGSPVGAALPGEFRALDEPRAVGVLSIGTIGLILLVANGLIEVDGILKEKA
RPGALSALGAIAPLAGAASALVADPGSPSAAGASGLARLARLVGIMRQRELRS
RSYAOVTLAGVSEVAANVLVAMTSSSYASAPALAAVARSALLASGLVNLVNGR
LTHLMRWVADATRVSKGVSVLVLTFLTALALORGLHPLGLFAGLVNLVNGR
NRPLIDGVOITLVAGLFAPVFYIAGMRDVADOLRPAWGVALLIATATAKVPA
LGARLGGIRGSEALVANGIMKGGCTDIATVIVVEGLLSNEATYMAVATVYVA
SPALILWEKRAPPTOEESARLERERARVITPEVERILVPIVAHALPGLADTVES
IVASKRKLGETVITELSVQDAPGPSAAGASGLARLARLVGIMRQRELRS
IOALIRASRDIDLVTIGARSPARAGMSGLQDAIVORASNVLVVVGDPAPAEAS
ARLILVPIGLEYSFAADLAHVALADAEVILLSAOTDPCGAVVMDRSPRAV
ARSVDAEAFRGRIGRVSSRVHGAHPSDEIRRELARVYDLVLVGCYDHPLGRL
YLGSTVESVYRSRPVALLIYAHGSTRQVR"
complement(5612. .5992)
/note="ORF4"
/codon_start=1
/transl_table=11

/product="unknown"
/protein_id="AAE26907.1"
/db_xref="GI:6724241"
/translation="MDKRICRCAIAEGYIPGSGNPEPOMTSHEACILNADRA
OATIVYFSDRDPAGPRVTVPARKTRHAFNDLITEPEPIPROTDIASVIESDPIV
QHRLDSRQAEMLSTIAYTRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE26908.1"
/db_xref="GI:6724242"
/translation="MKHVDGRFRFRICHTLGLASMLAGCGPSEKTVGSTRLP
GADARVADVPDAATTRLAVDVVHLSPPERLSEGFYVMQRPSPESWRRVGYLD
YNADSRGKLAETTVPANELLITAEKSSPSPSAATIGPISVC"
7610..11875
/gene="epoA"
7610..11875
/gene="epoA"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAE26919.1"
/db_xref="GI:6724253"
/translation="MADRIERAEDPIAIVGASCRPLGSGVINDLSCFWTLLEGSDRV
GLVEYCEALEENAIAPSAIVGTETGTCISYEALPQATASAEIDAHGSLGTM
PSVAGCISTALDRCPCVAVDTAYSSLAIVHACSLRSRGECSIALAGVCLMIS
STLVNLSKTRALARDCKAFAEADDFGEGECVAVVLRSGARADGRILLAVING
SAINHDGASSGLTVPNSSQSEIYLKRALDAGCAASSGVYEAHGCTGTGDPLEIDA
LNAVYGRDVAATPLIGVKTNGHEVAGSITGLKTVLSLOHGOIPHLAHOAL
PRISMGDLRTVTRARPMPTMNTOPRAGVSFGMSGTNAHVILSEAPATCPAPR
RRRELVLTSRTASALDAQARLRDHEITPSCDGVASLTATTSAMHRLAVAT
SRRELALDAAGOTSPGAVRSIADSSKGLAFETGGAGOTLGGRCGLVYWSAF
REAFDLCVRLNEOLDRPREVMAEBSADVALDQTAFTOPALTFEVALALWMS
MGVEPELVAGHSIGELVAVACVGFSEDAVFLVAAGRMLQALPAGAVSIEAPAS
DVAAVAAPHAASVIAVANAPDOVIVAGAGOPVHATAAAMARGARKALHVSFAFS
PLAMPLEAFRGVESVRSRPSIVLSNLSGRACIDVESPCGYVHAEVAFBAG
VKALHAAGACTPVEVGRKSTLLVLPACMDPARALLASSRAGDEPAIVLEAGCIGM
AVGLVSMALFPSSGGRVPLPTVPMQREMYWIDTADDDARDDRPGAGHVEEG
GAVRGDRRSARLDHPEPESGRREKVEAADRPFREIDIPGVLDHVLVYTRRAG
LGEVEIADVADAGLSFNDVOLALGMVDPDLKPNPRLLEGCEAGRIAVAGEVNCIY
VGOPVIALSAGAFATHTTSAALVLPPOALSIEAAMFVAVITAYALDRJARLOP
GERVLIHAATGGVGLAAVOAHVGAHVHATAGPEKRAVLESIGVRYVSDSDRV
ADVRAWTGGEGVYVNLSSGELIDKSFNLRSHGRVLEGKDCIADNOLGLRPELR
NLFSVLVDLKGMLERPARVRLLEEGLGIAAGVFTPPITATLPARVADARSMQ
AOHLGKLVITLGDPEVOIRIPTHAGAGPSTGDDLDRLASAPAAARAALEAFRLTQ
VSOVLRTPEIKVGAELFTRLGMDSLMAVELRRRIEASIKLSTTEFLSTSPNIALIA
QNLDLALATLSLEVAENLRAGVQNDVFSSGADQDWEITIAL"
7640..8920
/gene="epoA"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoA"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoA"
/note="Region: enoyl reductase"
11549..11764
/gene="epoA"
/note="Region: acyl carrier protein"
11872..16104
/gene="epoP"
11872..16104
/gene="epoP"
11872..16104
/gene="epoP"
/note="EPOS P: contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11

/product="nonribosomal peptide synthetase"
/protein_id="AAE26925.1"
/db_xref="GI:6724259"
/translation="MTINOLNLEHOGIKLADGERLQIOAPKNLNPILARISEH
KSTILLMLRRLPRAESIVPAPRAHAPFFLTDIOESTWLAGRTAFVPGIATRYED
CTDLDVPRLSRAFRKAVARRMDLRHAHTLPDMQVIEPKVADAEITDLGKLSRTEA
RLVSLDAMSHRIYDTERPPLVHVAVVRIDERQTRLVSLIDLINDLGSLSIFKDWL
SFYEDETSLPVELSYRDVYLALESRSKSEAQRMDYMKRIALPPPTLPMKAD
PSLTKELRFRHTOEWLPSDSMGRKKRVRGRLTPGVILAESEVIGKMSAPRETL
NITLFRPLVHPRVNDITGDETSMTVLDITDTPDKSFEORAKRIOBOLWEADHCDS
GIEVOREARAVLGIQKALFPVUTSALNOOVGYTSLORLGPTVTSOTPOLLDH
QLYEHGDLVLANDIVDGFPPOLDLDDMLEAVFVLRRLTEEMWGSQVRSCLPAQLE
ARASANTNALLSEHTLHGFAFARVQDLPQALVARSARTLVEELSRSSRLGATLR
EOGARNTLVAVMKEGMEQVAVVLAIVLSGAAVYIDDLPERIHYLIDGEEVLY
LQOPWLDGKLSWPGIOLRLVSAGVGGDDPMPPIOTPSLAIVYTSSTGTPK
GVMDHGRANVTLDINEREIGRGGVRLATSLSPDLSYVDFGLIAGGTIVPDA
SKLRPAHNAELTERKVIYVANSVPLMLRLVHFEGRSDLSARSLRLSLGDMIPV
GLPGLQALRPGVSVISLGATEASLWSIGYPRVNDLSMASIPYGRPLRNOTFHYLD
EALERPVPVPGQLYIGVGLALGYWRDEKTRKSFVHPETGERLYKTDGLRYLPD
GNIEMGRDNQIKLGRYVLEIEETLSHPNVDAVIVPGNDANKLLIAYVP
EGTRRAAEODDASTLTERIDARAHAAEADGLSDGEVOKRLAHGRDLDPKPVYD
TGODPREAGLDVYARRSVYTLFAPVPEVGRFTSCLSVPGCATLKPFPYSAG
STYPVOTIAYKSGRIEGVDEGFTYTHPEHRLKLSDBGIERGAYRONFVDFEAA
FNLFVGRIDALESLYSSSREFCLLAGMAQLMEQAPSCNIGYCPVGOENFEQVR
PYLDLHSDVYVHGMIGRVDPRQFQVCTLGODSSRRATTSRPAAPGRQHRADMLRD
FLRTKLPEVYVTFVVELDALPLTNSGKYDRKIDRKDKPSSRHSGHYAPDALEEI
LVAVVPEVGLLEVYGLQOSFVDDGASISIHVRKRSILQRLDREIAITELFQVNLGS
LASGLRSDRSIDQDRNMDRIVEVRKKRRRS"
16251..21749
/gene="epoB"
16251..21749
/gene="epoB"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAE26920.1"
/db_xref="GI:6724254"

Query Match 100.0%; Score 101; DB 1; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGCGAGATGGCCCTTACAGGCGCTGATCCATGACGCGGAGCGCTGATGATGCTGTGC 60
|||||
Db 36162 CGGCGAGATGGCCCTTACAGGCGCTGATCCATGACGCGGAGCGCTGATGATGCTGTGC 36221
|||||
Qy 61 TTGATGAGCAGACACGACCGCGCTTTCGCGGGGTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACGACCGCGCTTTCGCGGGGTGCTGGCA 36262
|||||

RESULT 3
ARI93029
LOCUS ARI93029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J, Madison., Molnar,I., Zirkle,R., Cyr,D, Dawn. and
Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: us 6346404-A 12-FEB-2002;
FEATURES location/Qualifiers
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25359 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 60
|||||
Db 36162 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 36262
|||||

RESULT 4
LOCUS AR199551 68750 bp DNA linear PAT 20-Apr-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 60
|||||
Db 36162 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 36262
|||||

RESULT 5
LOCUS AR199559 68750 bp DNA linear PAT 20-Apr-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 60
|||||
Db 36162 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 36262
|||||

|||||
Db 36162 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 36262
|||||

RESULT 6
LOCUS AR199567 68750 bp DNA linear PAT 20-Apr-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 60
|||||
Db 36162 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 36262
|||||

RESULT 7
LOCUS AR201097 68750 bp DNA linear PAT 20-Apr-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2,4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 60
|||||
Db 36162 CGGCGAGTGGCCGTTACAGGGCGGTGATCATGACCGGAGCGCTCGATGATGCTGTC 36221
|||||

QY 61 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 101
|||||
Db 36222 TTGATGAGCAGACACCGACCGCTTCTCGCGGCTGCTGGCA 36262
|||||

Db 36222 TTGATGACGACACCGACCGCTTCTCGCGGCTGTCGCA 36262

RESULT 8
AR208671
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn., and Goriach,J.
TITLE Genes for the biosynthesis of epothonolones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 60
|||||
Db 36162 CGCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 36221
|||||

Qy 61 TTGATGACGACACCGACCGCTTCTCGCGGCTGTCGCA 101
|||||
Db 36222 TTGATGACGACACCGACCGCTTCTCGCGGCTGTCGCA 36262
|||||

RESULT 9
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothonolones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source Location/Qualifiers
1..71989
/organism="unknown"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 60
|||||
Db 30550 CGCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 30609
|||||

Qy 61 TTGATGACGACACCGACCGCTTCTCGCGGCTGTCGCA 101
|||||
Db 30610 TTGATGACGACACCGACCGCTTCTCGCGGCTGTCGCA 30650
|||||

RESULT 10
AR166425
LOCUS AR166425 33529 bp DNA linear PAT 17-OCT-2001

DEFINITION Sequence 3 from patent US 6280999.
ACCESSION AR166425
VERSION AR166425.1 GI:16241741
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 33529)
AUTHORS Gustafsson,C., Beliaich,M.C., Ashley,G., Julien,B. and Ziermann,R.
TITLE Sorangium polyketide synthases and encoding DNA therefor
JOURNAL Patent: US 6280999-A 3 28-AUG-2001;
FEATURES
source Location/Qualifiers
1..33529
/organism="unknown"

BASE COUNT 4489 a 9518 c 14470 g 5046 t 6 others

ORIGIN

Query Match 51.1%; Score 51.6; DB 6; Length 33529;
Best Local Similarity 70.4%; Pred. No. 0.0014;
Matches 69; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 3 GCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 62
|||||
Db 14155 GCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 14214
|||||

Qy 63 GATGACGACACCGACCGCTTCTCGCGGCTGTCGCA 100
|||||
Db 14215 GCGGAGTGGCGCTTACAGAGGCGTGTATCCATGACGCGGAGCGGCTGATGATGCTGTC 14252
|||||

RESULT 11
AF319998
LOCUS AF319998/1 49736 bp DNA linear BCT 28-MAY-2001
DEFINITION Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
complete sequence.
ACCESSION AF319998
VERSION AF319998.1 GI:14210834
KEYWORDS
SOURCE Stigmatella aurantiaca.
ORGANISM Stigmatella aurantiaca.
REFERENCE 1 (bases 1 to 49736)
AUTHORS Silakowski,B., Nordsiek,G., Kunze,B., Blocker,H. and Muller,R.
TITLE Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium Stigmatella aurantiaca Sgals
Chem. Biol. 8 (1), 59-69 (2001)
JOURNAL
MEDLINE 21110452
PUBMED 11182319
REFERENCE 2 (bases 1 to 49736)
AUTHORS Silakowski,B., Nordsiek,G., Blocker,H. and Mueller,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig 38124, Germany
FEATURES
source Location/Qualifiers
1..49736
/organism="Stigmatella aurantiaca"
/db_xref="taxon:41"
complement(2..1306)
/note="ORF1"
/codon_start=1
/trans_table=1
/product="putative aspartokinase"
/protein_id="AAK57182.1"
/db_xref="GI:14210835"
/translation="MRPISARQAOGRGTSALALIVOKYGGTSVGDTERIKVNAKRCIA
AKACHDVVVVVSAMSGETNRLLKLKLVITEPNREDVDVVAATEQVSGIYALAIQ
TOKRAVDFIGHOVRITDPSFSKARISIDERVLDLKKKIVVAVGFOGODEHN
VTTLCRGSGDPTTAAVLAALAKADACEITDNGVYVTPDNNVPARXLERIAYEEMLE
IASVAKVLIQTSVDFPAKRYKPLWVKSSFTDPDTLICEEDANMEDIVLSIADKN
EAKIARIGVPDVPVAAKIFGALDEQSIIVLDIIVONVSKDGTDTVFTFVGRADLKAK
DVVKRIANSVKAEGVETDQVSKSVISVIGVMRNSHGVAAKMTYVLAGAVNNOMLISTS

gene
CDS
1363. .2895
/note="orf2"
/codon_start=1
/transl_table=1
/product="putative histidine ammonium lyase"
/protein_id="AAK57183.1"
/db_xref="GI:14210836"
/translation="MSRRLNIGDITLKLEIIOVAHHYTVELAPRAARVRAARDL
VDRAAGDTPSYGINTGFTLAERYIDKIDRLREIQRLIISHAGVSPILPEPARVL
LLRCNVLKANGYSIGIREPTELALLEMNRDVRVPPERGSGVAGSLAPLALHVLV
GEAEAFKGERLPAQAOLERAGLKPVLLEAKGIALVNGTOAMCAVGTLLQLRKEMA
DLADIAGMTLEGLGSKHPFPEIPIOVRAHEGOKACAHIRELLADSLAVEHVNCS
KODPSYLRGMPGVHGAAREGLSPARILVEYNISATDNPLEVETERIYSGAGNHG
KPSIADVAAMALFQLSAISERPYEOLVYNALSLPPLKNSGLNCGIMAOYTSAA
LVASRYLSHPASVDSTPSSAGREDHYSMGNTAALKROVAADPFRSCALIELVAQA
LDYRQPTRAKGPQAAVELIRSKIPTMEKDEIRHIDIAAVALIDSGEILNNAVTAAR
GQ"
complement(3275. .7822)
/gene="mxa"
complement(3275. .7822)
/gene="mxa"
/note="non-ribosomal peptide synthetase; contains a
condensing, adenylation, and PCP reductase domains"
/codon_start=1
/transl_table=1
/product="Mxa"
/protein_id="AAK57184.1"
/db_xref="GI:14210837"
/translation="MSYRELITTLAANGVRLLEGDKLQVASKAVLTPPELAELSAN
KADILALRKHSREETGSELQGLAPREYVPLSSGOERLWFIDRLSPGTSQVNHG
LRAGALDTRALRSIDELVLRHEVLRTHPEVSGSPROYLVSPDGTVALITVPLG
TESQREBEIYRSDLELAQOPEDLAKGPLRTVYALSPDPALFVTYKHIIITDWSIG
VFERESMLYFVRGQAALPVSQMPADSLAKERMLAGESGAREKRVNKGKGL
PPLQLPVDHNVSTSHRGATVPLALSADLARELASREGCEFFVLAARAFKALLH
RYSQVDPEGVTYANRGSVPADLIGFIANTLALRCDSGEPTFSQMLRAHKYVLA
LDYQALPFEVGVTCASRDGGLNPLVRACFTLEISIPAPTLIDPGTSMFLNKGPDSS
VEGAKFELSLILAASEKGLAMEYSREVDASTVERMWGHQVOLLSEIYAHPEVPL
SKPLTLAEBERGLINDMNGPVLDPYALCMHELVQAABERTPOAVAYVSGOKITTYAE
LNRRANOLAHHLRLGIGQEKERVLGCEKREDIYIGLAILKAGAVYVDPAPAKER
LAILLEDAQVPLTQORLVPELPATQARVACDKMDPTIGAESDINPEIRTAEBELA
YLIYTSGSTGPKVMTEHNAVAFILIMAMSVSPKPELAGLASTSCIPDLSEIET
PLCGAIVYIAKNALIELELPAAREVLTINTVPSAMGALLRSGSPSVIALYAGEA
LAGALVDOYIOLDHVRDVFNLVGPSETTITSTFVRNRCOPTTIGRPVGNVOYALDS
NREMPYIGVGEVYIGSGMYARGLGRPELTAERFVSPRGSGSEALIRKRGDLARML
PDGULEYLRADHVKLRFRTELGEIYAVLMEHSGLRDAVVVRREGKADKQVAVY
VGRGEKAPPEAELENDYLSKLPETMVPELTVGLDALPLTPNGKVDRAALPAERTHSG
PAKEHVAPRTGGEESLAIWROYLVGIGAHDFELGSHLLRYVILYARASGA
DIPLRALLOPTELEMAAVEAAKTGSLPAHDVYVEADAVLDAEIALGKALPVTG
ALRTIILTGANTGLGAFLLFELICRRTDARIYCIYRSKTBEGMNRIRKNIESYIANE
ALAPRIYVAGDITGQPLIGLSEKFORLSEIDITYNICALVNFLYIYISKRANVIG
TRELRLATRTRIKPLHYVSTVSVLPLGRKAPIREDEPLEGSSLVGGIYASKVAEK
LVREASRRGLPVTILRPGRYTGHSTRGAMWNTDLCYRCLKGCVMGVAPSYDALDIE
PVDYSSAIVDLRMRPESIGQTYHLVNPQEVRADEMNNYKRAFYGRLVPEYDQMLSE
LGSAASSDSEGLDLMLOOVPEPDRSVGGRBMVVCDSGDTLKALGGTCTSCPSVDAS
LISYISSLIVHGRGLKAPVR"
complement(7819. .11520)
/gene="mxa1"
complement(7819. .11520)
/gene="mxa1"
/note="PKs; ketosynthase; acyltransferase"
/codon_start=1
/transl_table=1
/product="Mxa1"
/protein_id="AAK57185.1"
/db_xref="GI:14210838"
/translation="MRAGAGHPLLGVSFTTSYOKENRFMESALSAKEPSTLADHYG
GAVVPGAAVLEMAISAARELSGKAVELVDTSPKEGILLPEGOERYVOMALSEGGD
ALSRQSSQKASTEPNAPAGMHAHYVQIIRTEGSPSTQIPRESEAIRERKSEVIAQA
AHYELASRGAIVGPAFGVOQIWRGAREALGHRLEPELRAARSARIRHALLDACF
QLVAAVAOEDLPPDAGPVAVALAHLRLHEQPGTEVCHCARLRSPSGARGGVYFEDL
VLRDASGRIVEALGRLVQODLADPAEHAGKDLFFLEMRVASSAVQOYVAGPSASGR
MLIADGSELADYVESQLRERHGEVIVINDLAGAGGRVVDATSPQAFPIGYLSAFAFG

gene
CDS
GSABRYIHLTGLPREVPDIMGCGVYLHVALRVSLSASPLMLVTRGVHHGSGS
RSASDVYQASILMIGIRTLIAHNSLKCIRYDVSAAMGGAALMBRELADDEEEL
SLRADGHVNGIIVGSPGRAPSEAVVAGCGRPFLVEDACORLEFRAGARRPPGCEL
EIEVEVALGGEGGTGPERNEGAGEBGIAGCGRRVLAIGEGRVNAVEGRATVAGSL
GSHVSPAGCTAIRPSSSHGEAALATVLPAYALFHLGRLEKRRVWVLAATSGL
GRAIKLAKRAGAEVFAVATDEORASIRLEGLDAQVDMSPASVPSRVLDATGSRGE
LAIYAPGAELEERSLPTLSECARVALLETILRTAPQEAALQOPREOLARVLRDSAR
FASIMOEKVAFEAGGLPSAAADKALKALDKPARKRVVPADARLADGTYITGGL
GGGLAYAKMVEGAGHLYLVGDDTYLSPEDQVYVTRALEAAGRIVYARADYDBRO
IARILSELBAGAPLREVTIHAAGVLDGVLQDVERRRRAKRVJGKNHVLTR
ALDPEFVLYSSAAMPFAGPGQNVVACNAFLDALAHHRALGLGLSINMPFSEVGL
AAQANGERLPAQCGSDSLTPAEENTILGRLLDGDVDMQAMPLELRKWFYRRAS
SPYSELVPAQOGEARVALLETILRTAPQEAALQOPREOLARVLRDSAR
IDPEALISGELDSLMGELERLNRIVSGIYVSLPSLIMKHPTLAALCAHLIGEMDRP
LAETLAQAACGAGVAYDESADDEKVEYVL"
complement(11487. .14303)
/gene="mxa2"
complement(11487. .14303)
/gene="mxa2"
/note="PKs; dehydratase; enoylreductase; ketoreductase;
ACP"
/codon_start=1
/transl_table=1
/product="Mxa2"
/protein_id="AAK57186.1"
/db_xref="GI:14210839"
/translation="MSKOEKAPDSALQORALLVEKQARLDAVERARTEPIAVTGL
GCRPPGAVDSSYWKILRDGVDALREVPESRMVPQGFIDSTRVPGKMGTRGFLD
DVEHDAEFEGISPREAASLDPOORLYLEVMAELENNAGLAPDOLASKGVGVAHS
SDYMARILKENDATRFDGYMATNGSYSPRVSRYVLAGLQPCMPVPTVATSSVLVH
LASLSRGESNLADGCVNLLSPMFCIKSGMOLASDRCSTPFAASVYVGRGEG
CGVILVIRKISNADGDGSLITALTGSVANHGAAGGLTVPPGSDQAOVALADNART
APALVGTIEANGTPTPGDPIELALNAGVLCGRKPRADRPFTIGSVKTNIGLBERACI
AGVTKTILISLOKREIPLHLRTNPHWMDITPARVVERVPPAHGRIATGVSSG
GLSGINAHVLEEAPSTAPTPVAEEDKQOLVLSATPRALSILAAWSSFLKINEG
GSLADISTSALESTPHEHRLALAKSKEOELIEDLEAFAPAKSHAGSLSGRGTGLQ
KIVVFPQGSQWIMGQORLYAEVNTFPEALIDMCEHAIISRTDMSRLDLNPPQOOR
WMDIDVIOPLTALOVLAAVYMRGLGEPHVAHVSHGEVAAVAGLSLEDAARI
CERKSLIRVSGCMAVYDLSRQAEYELIGADDRAIANSNRPSTVYSGDPAALIK
EYERLERREVFECWVRVAVASHSPQMDPLKODLDRNALIRTPGSAVPLSTVGAP
IDGELSAPYVNRILQRPVLEPADAVOURLIDGNVLELSHPILVPEVDMLEBGT
SGRLVLSPLRDDDEWPSLISVAGALFTVRIAMTGSQYIPTVAGLRFRTRSAGS
TGWSCQOOLRRAGSALAREVPS"
complement(14300. .29878)
/gene="mxa3"
complement(14300. .29878)
/gene="mxa3"
/note="PKs; modules 1,2 and 3 contain KS, AT, DH, KR, and
ACP regions"
/codon_start=1
/transl_table=1
/product="Mxa3"
/protein_id="AAK57187.1"
/db_xref="GI:14210840"
/translation="MANRNDNSELSPAKRMVLVALEKKOTRLNAVEGAKPEIAYDM
ACRPPQADNVFAFWRLDGGVDAVREIPLPEWPDVDAIDGMAKGRFLDSYSDG
PGFTGISPREAVRMDPOORLLELYAMELEBAGLDVDKLSGSRGVITGKNDYRCH
QVERPETGDASFATGVASVLSGRSLSTFNLQGSVLVYDTACSSLVSILHACOSLRN
RECMNALAGVNLILSPQSVLYSKLQALSDGSKAFDASANGFTIREGGGIVLKR
LSDALADGDLIATIRGSAINODKSGTGLTPPNVLSOQALROALESGLAEBOVYR
EAGHGTDIGDPIETAEALRETYGVPRGSGVSGSYTNGHLESAAGIGIKVYL
AMRHOISIPHLIKOVNPRLOESALRTIPLRLPWKRAVSCOPRRAVYSFGISCTNH
VITEBAPVAKDLPORPEVPLSANSRELVLAERYARLRRBEGGSLSDYCTRA
SAHRTHTHRLAVLADSEHTERLRAFAREOPSLAGSKARAGRRKTVFYVPGG
LAQVAGMGLAMESRPFEDALVVKDKALQPLGDSKAFDADQORLADIDVNOY
LFAVEVALTALMRSMGTEPDVAMGSHMGEEVAAVAGALSLEDAVLLIICHSKLMRL
GGRGATAMVLELVEEAEELIRPNWNGITMGGLNGPSROYLSGPEAVKEVLELGR
GVPRWYKQVMPSHRPLVEYVIRSELEGIAPIITPRADIPICSTVTEGPELDGALDA
HWGNLNRPVFAAAGHLLASDHGTPEVSPHYVLPAYDQYTSAGTBEVYASLR
HNDEERHTMPLSICALYSGHRVAMNRYVPGKGLVPLPTYPWORERWMPAPVAGV
ASRTSGHPLIGERITLISAOGEVRLMOVOYQPRHPLFLAHRGVPALPAAMAAM
ALSAKQTLIGVLEIDVFSALLSPFEDERVQLMISTERRPQASVFLFSAVME
TPSDLEIWMALHAGBSLRHGDASVVEKAPVAVLDELARCGSEASVEAHYQTLKEIGLE

188042	LOCUS	49377 bp	DNA	linear	PAT 10-AUG-1998
DEFINITION	Sequence 1 from patent US 5716849.				
ACCESSION	188042				
VERSION	188042.1	GI:3407982			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 49377)				
AUTHORS	Ligon,J.M., Schupp,T., Beck,J.J., Hill,D.S., Neff,S. and Ryals,J.A.				
TITLE	Genes for the biosynthesis of soraphen				
JOURNAL	Patent: US 5716849-A 1 10-FEB-1998;				
FEATURES	Location/Qualifiers				
Source	1..49377				
BASE COUNT	7247 a 19522 c 14477 g 8131 t				
ORIGIN	/organism="Unknown"				
Query Match	43.6% Score 44; DB 6; Length 49377;				
Best Local Similarity	65.0% Pred. No. 0.11;				
Matches	65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;				
OY	1 CGGCGAGTGGCCGCTTACAGGCGCTGATGCATGCAGCGGAGCGCTCGATGATGCTGTC 60				
Db	6277 CCGCGACTGCCCGCTTACAGCGCGGCTGTTTACACAGCGGAGGAGCTCGCAGCATGGCTGA 6336				
OY	61 TTGATGAGCAGACGACGACCGCTTCTCGCGGGTCTGCG 100				
Db	6337 TCGGCGACATGACGCGCGAGCGCATGCAGGCGGCTCTGCG 6376				
RESULT 15					
LOCUS	SCU24241				
DEFINITION	SCU24241 67523 bp DNA linear BCT 24-APR-2002				
KEYWORDS	Sorangium cellulosum acyl-CoA dehydrogenase (sord),				
ORGANISM	beta-hydroxybutyryl-CoA dehydrogenase (sord), putative				
KEYWORDS	methoxymalonyl-CoA synthase (sord), reductase (sord), soraphen				
ORGANISM	polyketide synthase A (sord), soraphen polyketide synthase B				
KEYWORDS	(sord), O-methyltransferase (sord), beta-mannanase, and				
ORGANISM	xylanase-arabinofuranosidase bifunctional enzyme genes, complete				
KEYWORDS	cds; and unknown genes.				
ORGANISM	U24241				
KEYWORDS	U24241.2 GI:13346872				
ORGANISM	Polyangium cellulosum.				
KEYWORDS	Polyangium cellulosum				
ORGANISM	Bacteria: Proteobacteria; delta subdivision: Myxobacteria;				
KEYWORDS	Mycococcales; Sorangineae; Polyangiaceae; Polyangium.				
ORGANISM	1 (bases 1 to 67523)				
ORGANISM	Beck,J.J.				
ORGANISM	Direct Submission				
ORGANISM	Submitted (07-APR-1995) Ciba-Geigy Corporation, 3054 Cornwallis				
ORGANISM	Road, Research Triangle Park, NC 27709, USA				
ORGANISM	2 (bases 52293 to 58655)				
ORGANISM	Schupp,T., Toupet,C., Cluzel,B., Neff,S., Hill,S., Beck,J.J. and				
ORGANISM	Ligon,J.M.				
ORGANISM	A Sorangium cellulosum (myxobacterium) gene cluster for the				
ORGANISM	biosynthesis of the macroide antibiotic soraphen A: cloning,				
ORGANISM	characterization, and homology to polyketide synthase genes from				
ORGANISM	actinomycetes				
ORGANISM	J. Bacteriol. 177 (13), 3673-3679 (1995)				
ORGANISM	95325306				
ORGANISM	7601830				
ORGANISM	3 (bases 1 to 67523)				
ORGANISM	Ligon,J., Hill,S., Beck,J., Zirkle,R., Molnar,I., Zawodny,J.,				
ORGANISM	Money,S. and Schupp,T.				
ORGANISM	Characterization of the biosynthetic gene cluster for the				
ORGANISM	antifungal polyketide soraphen A from Sorangium cellulosum so ce26				
ORGANISM	Gene 285 (1-2), 257-267 (2002)				
ORGANISM	22035375				
ORGANISM	12039053				
ORGANISM	4 (bases 1 to 67523)				

AUTHORS	TITLE	REMARK	COMMENT
Ligon, J.M., Hill, S., Beck, J., Zirkle, R., Molnar, I., Zawodny, J., Money, S. and Schupp, T.	Direct Submission Submitted (27-Nov-2000) Natural Product Genetics, Syngenta, 3054 Cornwallis Road, Research Triangle Park, NC 27709, USA	Sequence update by submitter	On Mar 15, 2001 this sequence version replaced gi:868150.
FEATURES	location/Qualifiers		
source	1..67523		
	/organism="Polyangium celulosum"		
	/strain="So ce26"		
	/db_xref="taxon:56"		
	57..986		
	/note="hypothetical protein; Orf1"		
	/codon_start=1		
	/evidence=not_experimental		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AKI9886.1"		
	/db_xref="GI:13346877"		
	/translation="MRADNPSSAGICDAIRYIEHGOVLAAGRTYLFHYLEHGVSRA LFRAAEAFAVEEVEGTRAAFPASIGESKSPSPAQNAAPASSLDPIPAVDEPL NDELICVRDCVAATSQGCERLAEMVHGOPLLSOEWHMENAELAOCCRLDGTFFH ADLRANGVDPSMAAPAVPNTGSAACCTEPLPPRQOAAQORPAPHSSEDT HCPGKILLAVARKKEGVMIGSYGAHRLRRVDEIKTWPEDVOEAVLQHFWSIAG KSNAPGPISDGLESNPVPVPAARTTRPKMMRR"		
	1424..6127		
	/note="hypothetical protein; Orf2"		
	/codon_start=1		
	/evidence=not_experimental		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AKI9887.1"		
	/db_xref="GI:13346878"		
	/translation="MRSPFRNFNPSPPTWAGPRAATFHRDEDAFOQLVDGALREFGA GVAMPPEGRDGGIDVLENDARLPEFSRDLAPVIECDNDNDARGAENVAARK KVEEKILAKOSAGMGPTGPRMATARTLYLTSAALPMQDARNRLOHGOEFOFLRSD GRTTEHPVAVDWDCSLIDRNPRLADPRLGRLTSGOAGLASIGSVRIYNDL PTIAPGRESQTHPDRILAHYLERAGEDRGVLLVAGAGVCTRTILLEGQADAGNAV LHVASGEPPLSSSDLDQEVVDGRDVTYICDYDQRLADVVSIRQLPPAROGKRV AIIASARTPRDARSRNPDFFAVRRLALETGOALAGTMEQOIAPOAVMLGEKVR KLGDQPMIALVLTLELERLAATRKTLRADQTVRPEGLRHMLERLAEDEIVADGS SRMDEEATSHVLCAAALACFPARKPVAVITCVIAEENKAKAADAERTIVONIEL GMIRNAGKITVAHDVYTDQLDQALTTSDAVRNHRLRLRLPCKASPVLARFARA LRRYIAGGFPFKRELERRAQAWMLKASPALGRMAGASAKSSAALAMAMGSEWKT TTCQWDQIVAPVLRHGASDAMPPLVALICIRNLPBPVYARLEAIAVIEOHNEDE HAASILMLDLDRPDLGSAISRLFLGMLPMPEYEHRRQSPALVYGLLETEREAPALI ERLGGPTLRWMIHPEVEEAGLPLGMLREEDTSSALOSVEGSMWELRRHDAQTA VLDHAGLLTLCDEBSAHRVAADALPMKTIHMDTAGCIVGLTILRGSTIEQILPA VLDANAMFKRRRSRFSVFTILGELTSLKPSLDSDRYQOAKIALGWYAADRTPOART LVHVSRRGDLDEPRAGADIDREACSLIQKPRSDSLILGMLERTLDDVDAARA ELALGWLEKNVGEELHVLVLTASLACHVAVTPARLAADAIAQESGAVRNVDGVL GSLIIRGIETSLIORASAOAMQFDDHADLAIOAYRLILGRPDIDPMARLVG AOLMLRQHGMLGCAVAYIAPLLAREDTLPRLARETTATITMLEONHPREADVITNV LATSPRINKPGISVYAAAREWLOMSRRPLASLIANTLPPSAEPASPDIOHALSM LEAHGDPDAFDLLRLSARRDLSPRSASVYVSTGLSWAHPHAELAPLITLLVR PEIDASVYQVLSRAIDMDTTHELSESGYALAAVMAQRRRELAAVSGRPLGAV EMLRDASHSTIAGVITALLYSRKNLTPTTLAATATSVKMOGSHAHEESLWPLAHL QHNRLPGLARVYOKAVEWIDHHTHVGDAVYALHELFGDLDEDFRTVFLAALRL GHNVDLPEGQOOLRIVSOASPESEGODEGVESLVDTLASSVEGEPEDCARLP ALLVEARSRAPRLERRARLVYALRLDRRMHMLRRMAGACQRLLEADAWDRARA LQLLLEGLALPALED"		
	complement(7505..8896)		
	/gene="sore"		
	/complement(7505..8896)		
	/gene="sore"		
	/codon_start=1		
	/transl_table=11		
	/product="acyl-CoA dehydrogenase"		
	/protein_id="AKI9892.1"		
	/db_xref="GI:13346883"		
	/translation="MPLSLAPHEHARVERVOGVDEHVAFPADAFHRAORTPELTL LAERGVIGISIPKEFGAGEDAYVTILGLAAELARGCSVNSLTVHFMVALILRNGT		

